A Guide to

MODERN

J. K. MEHTA
(RETD.) PROFESSOR OF ECONOMICS
UNIVERSITY OF ALLAHABAD

MAHESH CHAND
READER IN ECONOMICS
UNIVERSITY OF ALLAHABAD

ECONOMICS

SOMAIYA PUBLICATIONS PVT LTD □ BOMBAY

Jamshed Kaikhusroo **Mehta** (1901)
Mahesh **Chand** (1918)

PRINTED BY S. ABRIL, S. J. AT THE ANAND PRESS, GAMDI–ANAND, GUJARAT AND PUBLISHED BY K. R. SAMANT FOR SOMAIYA PUBLICATIONS PVT LTD, BOMBAY 14 DD

To the memory of

GURU DATT KARWAL

who taught us economics
and who initiated us
into the intricacies of
ancient Indian religious thought

Preface

AN ATTEMPT is made in this book to introduce the student to some of the most significant parts and features of modern economics. The tools and techniques of analysis now used in the science of economics, the various features of the science that have come more prominently to the forefront in recent years and a number of concepts that have enriched economic discussion and led to a clear understanding of the problems and principles of our science, make up the contents of this book.

The coverage of the book, as can be seen from the Contents, is wide. The treatment of the topics taken up is quite comprehensive without being very detailed and exhaustive. It has been our object not to overload the book with details that a student can easily find in the literature. The chief merit of the book, if any could be claimed, is that it saves the student much otherwise unavoidable time and labour in going through the literature to collect the material that lies widely scattered.

Care has been taken to place before the student every topic, every phenomenon, in its natural setting, with its legitimate relationship to allied phenomena clearly indicated. For example, while writing on *effects* in economics we begin with the logical and scientific meanings of the words cause and effect and pass on to a brief mention of *effects* in physics before taking up the *effects* in economics. While explaining the Ricardo *effect*, a description of cycles, turning points and time-lags is given and the causes of the trade cycle are briefly discussed. Here savings and investment come into the picture. An explanation of

the Pigou *effect* has necessitated the discussion of the meaning of employment and the causes of unemployment and underemployment equilibrium. *Effects* are shown to be due to the love of inertia. Discussion of Sigma *effects* has led to the posing of the question: Why not a Harrod *effect?* The importance of acceleration, its technical and psychological aspects, and how *o* and *g* differ, all come up for discussion. The Kalecki *effect* is critically discussed and checks to the increase of investment mentioned.

It is not necessary to multiply such examples. The Table of Contents of the book will give some idea of the matter that is packed into each chapter. Our individual contributions are indicated in the Table. Along with other material, the more mathematical developments in modern economics such as the Theory of Games, Input-output Analysis, Linear Programming, Aggregation and Social Accounting, have been brought together for the benefit of the student.

We take this opportunity to acknowledge our debt to those whose works we have drawn upon even where they have not been referred to in footnotes.

We are grateful to our colleagues, Sri Ram Narain Lohkar and Sri Dinesh Kumar Srivastava for preparing the Index and for the help rendered in reading the proofs.

It is hoped that the book will prove of much use to undergraduate students and of some help also to more advanced students.

J. K. MEHTA

MAHESH CHAND

Contents

PART TWO: TECHNIQUES OF ANALYSIS

PART FOUR: MODERN CONCEPTS IN ECONOMICS

PART ONE

Mathematical Tools of Analysis

CHAPTER 1

Modern Economics

Meaning of Modern Economics

THE TERM modern economics has no precise significance. Like many other words, it can have either a historical or an analytical connotation. Thus, modern economics might mean either the body of knowledge that is historically speaking recent or that which is analytically new. The views of some old economists are, in the analytical sense, modern though, in the historical sense, they are not. It is the characteristic of great thinkers that, though they become historically old, they continue to be analytically modern. Great minds think always far ahead of their times.

A student of economics generally understands by modern economics the stuff he finds in modern books. He is concerned with the study of those books and it is natural for him to attach to modern economics the meaning he actually does. The economics that modern books contain is, in the main, modern in the analytical sense. We find new theories as well as old ones stated in a new way — old doctrines explained and examined with the help of modern analytical tools. However, when the tools are changed the doctrines themselves undergo some modification, even where they constitute a mere refinement of old ideas.

Modern Tools of Analysis

We have two tools of analysis which have been applied to economic

theory since its inception. They are logic and mathematics: the former is concerned with the laws of correct thinking; the latter with the application of these laws to quantitative relationships of economic entities. These are not definitions of logic and mathematics; they merely indicate what they are concerned with. Today we apply the same tools of analysis to the solution of problems in economic theory. But the use of these tools has been widened and deepened. We now cast in mathematical mould the economic propositions which, in the past, were only verbally stated, and we use higher mathematics to develop our theories and bring out their full potentialities. One cannot say, perhaps, that the use of logic has also undergone a change. But mathematics is itself the logic of numbers and so, with the increased application of mathematics, the laws of correct thinking have automatically found a greater use for themselves in our science.

Now that mathematics is used on an extended scale, an average student of economics finds it difficult to understand the new language of economics. One needs to be a mathematician to follow modern developments in economics. Geometry, algebra and calculus now play an important part in the solution of all economic problems. Thinking has become, therefore, more rigorous and the nature of relationships between economic entities has become very precise. We have been able to discover unsuspected relationships and express the same in more accurate (quantitative) language. Economics is fast acquiring the characteristics of the natural sciences.

There is another tool which we have lately begun to use. That is the electronic computer. It is a physical tool though, like every physical tool, it is constructed and used with the help of the human mind. The more mathematical economics becomes, the more it lends itself to the use of machines. An electronic calculator is, in some respects, very much like the human mind but it is capable of performing its task much more quickly. This is because the mind of a man is not purely mechanical in nature. It is open to emotional influences that springs from the heart, as it were. Due to this interference from an outside source, our mind works slowly and is not able to retain and interlink the various results of its operations.

The process of development in economic theory is from the verbal plane to the numerical (and pictorial) plane and from there to the mechanical plane. Our thoughts first mould themselves into concepts expressed in words. Thereafter the magnitudinal aspects of those concepts, or their interrelationships, are expressed in terms of mathematical symbols. In the end, the operations needed to use these symbols for the purpose of our theory are entrusted to mechanical tools. All these steps enable us to transfer the work of acquiring knowledge from the mind to an external agency governed and controlled by it.

A machine not only relieves human beings of physical toil; it also relieves them of mental strain. But it may be that we are tending to overstep the mark with the result that, instead of reducing the strain on our mind, the use of complicated machines greatly increases it. Our life becomes complicated and what we gain in one direction we tend to lose in another. However that may be, we economists now have in our hands more complicated and more automatic tools for the formulation and refinement of our theories than our predecessors could even dream of.

Modern Technique of Analysis

It is possible to make a distinction between *tools* and *techniques*: whether it is desirable to do so is another question. For instance, we can use the *tool* of mathematics and adopt the statical *technique*. There are statical and dynamical *analyses* and there are macroeconomic and microeconomic *techniques*. Economists in the past employed the statical technique or analysis whereas modern economists are making an increasing use of dynamical technique. What these are will be seen later, but it might be mentioned here that in the former analysis we do not take account of the functional relationship between variables belonging to different moments of time, while in the latter we do. The statical analysis is, therefore, called timeless analysis. Dynamical analysis treats time as a variable. That expression is perhaps not a very happy one but it helps one to see that in dynamical analysis time makes an appearance. We can say that dynamical technique is a modern technique.

Another distinction is that between macroeconomics and microeconomics. The first treats of the economy as a whole while the latter studies it in parts. The first is, in a sense, a closed system while the latter is not. In macroeconomics we ignore the interrelationship of the various operating units. In microeconomics we concentrate on such relationships. Keynes was the first modern economist to emphasise the necessity of making a macro-study of our problems. He pointed out that what was true of a part was not necessarily true of the whole. If, for example, one employer were to lower wages he would be able to employ more men. But that is not always true of the whole economy. For, if all employers were to lower wage rates, the purchasing power in the hands of buyers would be reduced and there would, therefore, be less demand for their products. Employment, in such a case, cannot increase. This argument is not always correct. But here we are not concerned with the relationship between wages and employment; our object is to show that what is true of a part is not *necessarily* true of the whole.

Keynesian economics is variously described. It is sometimes called new economics and sometimes called macroeconomics. Further, some

think that it is macro-dynamic while others think that it is a macro-static theory. We shall have something to say about this in its proper place. Here we are concerned with the fact that there are certain techniques which we might call modern in the sense that they are being increasingly made use of in economics today.

Modern Concepts

Modern economics makes use of certain new concepts. In a way, there is nothing new under the sun; one concept evolves out of another. It is is not possible to have an idea that has no ancestry, as it were. Yet it can be said, in a way, that the economists of today make frequent use of certain new concepts. A student of economics should familiarise himself with the meaning of these concepts if he is to understand modern theory. In the pages of this book an attempt is made to explain those concepts in a simple way. Here, however, we shall simply mention some of them.

We have already spoken of statical and dynamical concepts and macroeconomics and microeconomics. We can have combinations of these giving us the compound concepts of macro-statics and macro-dynamics and micro-statics and micro-dynamics. These terms are now frequently used and so we shall explain them in the appropriate place.

Then, there are many mathematical concepts such as curvature, convexity and concavity; the gradient, positive and negative inclination (of a curve), functional relationship, coefficients, variables and parameters, derived functions or differentials, integrals, asymptotic rise or exponential rise of a magnitude, etc. These too will be explained. Other concepts are maximising behaviour, homoeostasis, and the balance-sheet approach, which are used to explain consumer's and producer's behaviour. Growth and fluctuations, programming, input-output analysis, indifference curve and indifference surface, closed and open systems, economic models, the multiplier and the accelerator, ceilings and floors, exogenous and endogenous forces, monopolistic competition, expectation, uncertainty, potential surprise, underdeveloped systems, disguised unemployment, national income, social accounting and functional finance are some of the other concepts, the precise meaning of which is not always clear to students. Then, there are other concepts such as the sigma effect, the Ricardo effect, the Kalecki effect and the Pigou effect. A student comes across these terms now and then. Unless he knows what they imply he cannot follow his texts.

Modern economics abounds in such concepts. Each of these has a special meaning; the difficulty of a student consists in the fact that his notions of these terms are vague. The terms used in economics do not have the same meaning as they have in common parlance. For instance,

we use the words flow-concept and stock-concept in relation to money. In what sense is money a flow? One who has seen water flowing down an inclined surface cannot imagine how money can be thought of as a flow. We shall explain all these terms as they constitute the raw material of modern economics.

Main Features of Modern Economics

If we were asked to single out some special features of modern economics we would mention the part that national income and employment play in the study of economic theory. Recently, economics has begun to be studied from the point of view of employment and income. Income depends on employment and employment on income. In such a study, it is clear, knowledge is sought with a purpose, the purpose being increase of social welfare. Employment and national income are indices of social welfare. The economic system is conceived of as a means to an end and it is studied with the desire to determine the way in which we can best help the system to achieve for us the ends we hold dear.

Economic theory thus treated as a means to social ends must run in terms of what are called aggregates of a system. We have necessarily to to make economics a macro-study. Classical and neo-classical economists were concerned mainly with the determination of the value of individual units. Modern economists are more interested in the determination of the value of aggregates of individual units. We want to know, not what a worker earns, but what a whole society earns. We want to know not how the price of a commodity is determined but how the general level of prices is determined. The present study of economics points to the shift of emphasis from the individual to the social aspect of all economic phenomena.

There is another direction in which modern study is tending to turn. That is the direction of *growth*: we are interested today in determining the factors on which the growth of an economy depends. The classical economists too had something to say on this point but the neo-classical economists, in the main, they with their followers, were busy studying the forces that brought about equilibrium in a system. They perhaps felt that what was needed was to facilitate the full employment of existing resources so that production could be maximised. The emphasis implicit here was on the concept of existing resources. What was needed was optimum allocation of resources to needs. Recently, economists have begun to pay attention to the forces on which the growth of factors of production and of production itself depends. Our concern is not only with the optimum use of given resources but also with the optimum way of increasing resources.

The shift of interest is from stability to growth. Thus, that from the individual to society and that from stability to growth or from static equilibrium to dynamic equilibrium mark themselves out as special features of modern economics. These can be mentioned as broad features. But there are certain other directions in which economists today are projecting their thoughts. There is the theory of games, for example, which has been applied to economic problems. Here microeconomics engages our attention once again. Individual producers compete among themselves and there is a certain pattern visible in their competition. Like chess players they follow a certain strategy with a view to securing the best results. They anticipate the reactions of rivals to their own decisions and on such possible reactions they base their decisions. The theory of games is a mathematical technique and one needs to possess mathematical equipment to understand it.

Coming back to macroeconomics, we have programming, general activity analysis and input-output analysis which are closely associated techniques adopted for the study of a whole economy through the interrelationships of its constituent parts. These techniques are mathematical and statistical in their content. Such developments in the science of economics are tending to divide economists into two groups: one that is mathematical and statistical, the other that is not. However, with the growing application of mathematics to the science of economics, the first group is attracting larger and larger numbers of adherents.

There are other features of modern economics which deserve mention. First, there is the growing interest in the study of problems of social welfare. An almost separate branch of economics called welfare economics appears to be developing. But there can be nothing like illfare economics; all economic studies are those of welfare. Hence, the so-called welfare economics is in reality social-welfare economics. The change is, thus, from the individual to the society. We have now what is called new welfare-economics and a number of books have been written on this topic. But social welfare is yet (and will for ever remain) a vague, abstract concept. One does not know who it is that is really interested in a social welfare which, at least immediately, differs from one's own interest. We shall have something to say later on this point also. But today welfare economics has only theoretical interest for us. No study of this side of our science gives us competence to tackle practical problems. But theoretical economists are not interested in the practical side of life. It is their object to study theory for its own sake, for the light that illuminates their way and not the fruits that it might yield. All such studies that hold out no promise of fruits are highly interesting to economic theorists and they serve the one very important purpose of sharpening their intelligence—the tool with which they weave their economic fabric.

Then, there is imperfect competition or impure competition which

Professor Chamberlin calls monopolistic competition that engages our attention today. Pure or perfect competition is the norm and deviations from it is the actual state of affairs. That which *is* is imperfect and impure; that which it is *tending to be* is perfect and pure. The study of economics has proceeded from the norm to the actual. That is the direction in which scientific study should travel. We must see the actual in the light of the ideal; that which exists must be understood as a deviation from the ideal. Thus we are increasingly getting to recognise the fact that competition is not perfect and not pure in this world, the world that we are interested in studying. But competition exists among like units. Producers compete among themselves and consumers compete among themselves. The recognition of the imperfection of such competition forces upon us the necessity of taking account of inter-unit relationships and thus making a micro-study of economics. In all such studies we pass on to the whole through its parts.

Next we have another feature of modern economics that shows itself in the growing interest taken in the part that the government of a country can play in directing and determining the current of economic life. The role assigned to the state in modern times is much wider than it was expected to play in the past. Public finance has ceased to be what it was when taxes were levied to obtain the necessary funds for the performance of collective services for the people. It is now tending to be functional in nature. What this means we shall see later on. And with this we have begun to recognise the importance of controlled economies. Our study of economics now has definite policy implications. It has been gradually casting off its positive garment; but we still have some who have stuck to their guns and for them economics continues to be purely positive in its content.

CHAPTER 2

Curves

Why Mathematics?

We shall not define mathematics for we are not writing a book on mathematics. It is sufficient for our purpose to know that mathematics is concerned with the quantitative aspect of entities. The basic measure of quantity is expressed in terms of numbers. But from numbers, which are abstract, we can pass on to pictorial devices and represent quantities or magnitudes by lines, areas and volumes. There is a difference between quantity and magnitude; there are no synonyms in the English language. But we do not need to be exact in the use of these terms; simplicity demands that we use these terms as interchangeable.

We can have no science unless the entities we are concerned with are quantitatively distinguishable from one another. And this distinguishability is the essence of measurement. The common usage of the word measurement is crude and unsuitable for purposes of scientific study of a subject. For us, therefore, measurement has a different connotation. A thing is said to be measurable if it can be classed as greater or less than another thing. A layman calls this comparison but we know that, in a sense, every comparison is measurement. To measure one thing in terms of another is to compare the two things in respect of some particular aspect. We can compare things in several ways and also express the result of comparison in various ways. Accordingly, we get different categories or types of measurement. Some are suitable for certain

purposes while others are not. If we like we can say that there are degrees of measurability or intensities of measurement.

Mathematicians speak of measurement correct up to monotonic transformation and measurement correct up to linear transformation (with an additive or multiplicative constant). With these we are not here concerned. All that we need to note at this stage is that mathematics starts with the concept of measurement. In economics we have to measure utility, supply, demand, value, and in fact everything else with which economics is concerned. If we have to measure these things and then establish relationships between one measurement and another we need to employ mathematical tools. We use numbers and other symbols to express quantitative relationships. We use arithmetic, algebra and geometry. We also use calculus, differential and integral. A student of economics must know all about simple curves and a little about calculus.

Curves

A curve is a line which does not project itself in the same direction. A line that keeps its direction unchanged is called a straight line. Such a line is also described in another way: a straight line is the shortest distance between two points (on any given plane). A line, we know, is made up of points. It is difficult to give a very logical definition of a point. Let us say that a point is the smallest possible part of a line. Go on shortening the length of a line and, in the limit, you will get a point.

Sometimes it is said that a point is that which has position but no magnitude. That way of describing a point imposes a great strain on our imagination. For, anything that has position, has it in space. It must have a location in space. And if it has such a location it must have magnitude; without magnitude it cannot occupy space and, consequently, cannot have location in space.

A line has no breadth, it has only length. Reversing the statements made above we can say that a point stretched out in length becomes a line. Or we can say that when a point moves along a path it traces out a line. When the point moves in the same direction we get a straight line but when it moves in different directions we get a line that is not straight; it may then be a curved or a crooked line.

When a line moves over a plane it traces out a surface. If it is a straight line and moves along the direction of its length it does not trace out a surface. If it moves at an angle to the direction of its length it does form a surface. A surface, therefore, has another dimension also. While a line has only length, a surface has breadth also. And we might note in passing that when a surface moves in a direction, not of its length or its

breadth, it gives rise to a volume—volume which has three dimensions, namely, length, breadth and thickness. Diagram 2.1 illustrates lines, straight and curved, and surfaces.

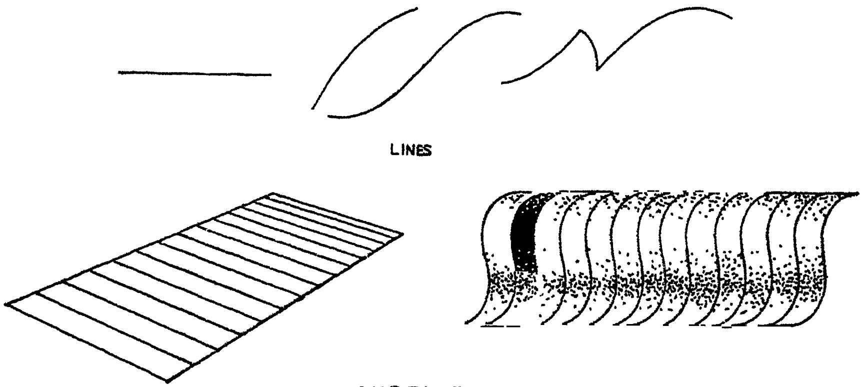


DIAGRAM 2.1

CURVES SHOWING RELATIONSHIP BETWEEN TWO VARIABLES

Let us begin outright with a curve drawn on the *X*-*Y* plane. What does this mean? We can represent the direction of length by *X*, say, and the direction of breadth by *Y*. In Diagrams 2.2(*a*) and 2.2(*b*) — we have two straight lines at right angle to each other. The vertical line is always called the *Y*-line or the *Y*-axis and the horizontal the *X*-axis. As we move along the *X*-axis we trace out length and as we move along the *Y*-axis we trace out breadth. With the two directions we get a surface. Hence the two axes give us a surface or a plane which we call the *X*-*Y* plane.

In Diagram 2.2(*a*) the curve *D* is drawn with reference to the two axes of *X* and *Y*. Every point on it shows its distances along the *X*-axis and the *Y*-axis. Thus the point *P* shows that it has travelled along the *X*-axis a distance *Oa* and along the *Y*-axis a distance *Ob*. When the point *P* moves down the curve it travels farther along the *X*-axis but backward along the *Y*-axis. That is why its distance measured away from the *Y*-axis increases while that measured away from the *X*-axis decreases.

The curve *D* is a continuous curve while the *ABCD* curve in Diagram 2.2(*b*) is a discontinuous curve. It travels from *A* to *B* and then ceases to travel from *B* to *C*. What does this mean? Take the curve *D*; every point has two co-ordinates, i.e., it has its distance from *Y*-axis which we

call the *X*-coordinate and its distance from the *X*-axis which we call the *Y*-coordinate. The curve *D* is continuous, meaning that there is a known value of *Y* for every known value of *X*. But in the case of the curve *ABCD* this is not so. When we pass from *B* to *C* there is a jump, as it were, from *B* to *C*. There are no known values of *X* and *Y* between *B* and *C* on the curve. At point *C* the curve starts from the same *Y*-coordinate that it had at point *B* but not at the same *X*-coordinate. Because the curve jumps from one point to another this discontinuity is called sometimes a 'jump discontinuity'. We shall see that when a demand curve has a kink in it the marginal revenue curve shows such a discontinuity.

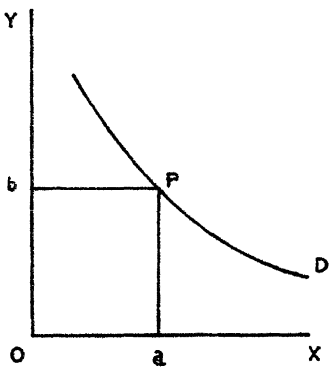


DIAGRAM 2.2 (*a*)

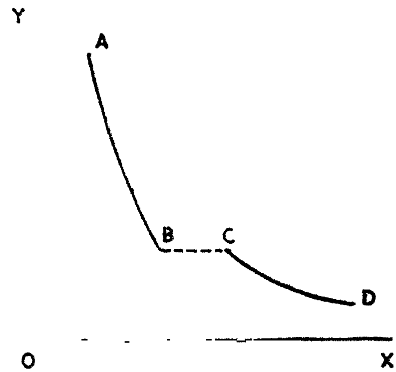


DIAGRAM 2.2 (*b*)

Taking the curve *D* in Diagram 2.2(*a*) we can say that it shows a certain relationship between the *X*-coordinate and the *Y*-coordinate. If we measure, let us say, price along the *Y*-axis and the quantity of a commodity bought or demanded along the *X*-axis, a point like *P* would show the relationship between price and demand. In the above diagram it shows that at price *Ob* (which is equal to *Pa*) the demand is *Oa* (which is equal to *bP*). The curve then can be called a demand curve. *X* and *Y* or, in this example, demand and price are the two variables. A change in one is associated with a change in the other. This relationship is indicated pictorially by the curve *D*. A curve, we can therefore say, is meant to show or represent a relationship between two variables.

In economics we study, as we observed earlier, the relationship that exists between or among the various variables of an economic system. When we want to show the relationship between two variables we can make use of a curve. We then get demand curves, as above, or supply curves or indifference curves which will be explained later.

Shape and Position of a Curve

A curve can be either positively inclined or negatively inclined. It can be convex or concave when looked at from the origin (*O*). What do these terms mean? In economics we often use curves whose inclination and curvature have economic significance.

The curve *D* is said to be negatively inclined for the reason that if we draw a tangent to it at any point such as *P* the trigonometrical tangent of the angle it would make with the axis of *X* would be negative. But this necessitates a knowledge (very elementary though) of trigonometry. Hence let us define negative inclination of the curve by saying that the curve *D* is negatively inclined because an increase of one coordinate is accompanied by a decrease of the other. It will be seen from Diagram 2.2(*a*) that as the *X*-coordinate increases (demand increases) the *Y*-coordinate (price) decreases, and vice versa. A positively inclined curve is one in which both the coordinates increase or decrease together. In economics, a demand curve, in all normal cases, is negatively inclined whereas a supply curve is positively inclined. A negatively inclined curve is a falling curve while a positively inclined curve is a rising curve.

The curve *D* in the above diagram is said to be convex to the origin because it bulges towards the origin. The property of a convex curve is that as *Y* (which means the *Y*-coordinate) decreases *X* increases by greater and greater amounts. This is shown in Diagram 2.3(*a*). A

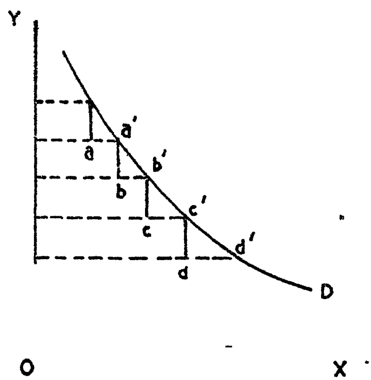


Diagram 2.3 (*a*)

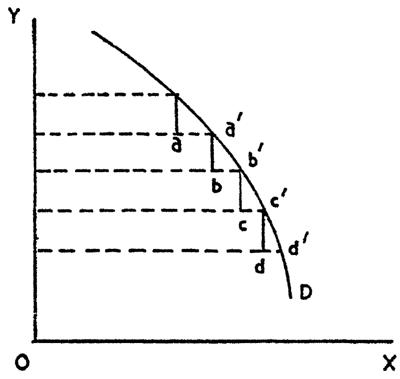


Diagram 2.3 (*b*)

concave curve shows that as *Y* decreases *X* increases by smaller and smaller amounts.

In Diagram 2.3(*a*) the *D* curve is convex and it shows that every time *Y* decreases *X* increases by greater and greater amounts. It is seen that

bb' is greater than aa' and cc' is greater than bb' and so on. In Diagram 2.3(b) the D curve is concave and it shows that every time Y decreases X increases by smaller and smaller amounts. It will be seen that bb' is smaller than aa' and cc' is smaller than bb' and so on.

When demand increases more and more rapidly with each fall of price, the demand curve is, therefore, convex and when it increases less and less rapidly it is concave. But of two curves, both of which are convex (or concave), one may show greater increase of demand with fall of price than another. It should not be concluded from this that elasticity of demand can be judged merely from the slope of a demand curve. We have simply explained here the meanings of the terms positively-inclined and negatively-inclined and of convex and concave curves.

ADVANTAGES OF CURVES

Before we give certain examples of application of curves, it is worth a mention that when we explain ourselves in spoken words, they have the disadvantage of vanishing as soon as they are uttered. The only place where they continue to exist is in the mind, but that is a different type of existence. We can also explain ourselves with the help of curves (in a diagram). This has the obvious advantage that what is communicated to others does not vanish as words do: the pictorial explanation remains there for the eyes to see and the mind to understand. What is more important, however, is that we can say, as it were, so many things in the same breath and others can see and take mental note of multiple things at one and the same time. The relationships that exist between variables can, therefore, be understood easily as the diagram offers a comprehensive picture of the system of forces. But the weakness of diagrammatical representation consists in the fact that, at best, it can indicate the relationship that exists among three variables of a system. This is because a diagram is a picture drawn in space to exhibit magnitudinal relationships. And space has only three dimensions. There are, however, many problems in which we need to know at a time the relationship between two variables only.

THE CONCEPT OF ELASTICITY

Solution of economic problems depends, among other things, on the elasticity of variables. Thus, elasticity of demand, elasticity of supply and elasticity of marginal revenue, all play their part in the determination of price. Besides the elasticity of these variables we are concerned in economics with elasticity of substitution and elasticity of derived

demand. We make use of these concepts in dealing with problems of the relative share of a factor in total earnings or in determining the effect of a change in the supply or demand of one factor on the earnings of other factors. A student can follow the solution of these problems only if he knows something about elasticity.

One who is not initiated to the mysteries, as it were, of elasticity is apt to judge the elasticity of a curve (which shows the behaviour of a variable) from its shape. Let us, therefore, explain the way in which elasticity is measured. Without the use of mathematics, i.e., without algebra or geometry, one cannot be precise in one's thoughts and expressions. The use of mathematics forces upon one the necessity of thinking in precise quantitative language. An expression of an economic principle in simple prose may lend artistic beauty to it but it robs it of scientific accuracy.

We start with Marshall's definition of the measure of elasticity of demand. According to him elasticity of demand is measured by the proportional change in demand divided by proportional change in price. If demand increases by 10 per cent due to a fall of price of 1 per cent, elasticity of demand is said to be 10. And for purposes of such calculations we have to observe the change in demand caused by a very small change in price. Mathematically speaking, price must be allowed to vary infinitesimally. Elasticity of demand is, thus, point-elasticity of demand. Geometrically it is measured as shown in Diagrams 2.4(*a*) and 2.4(*b*).

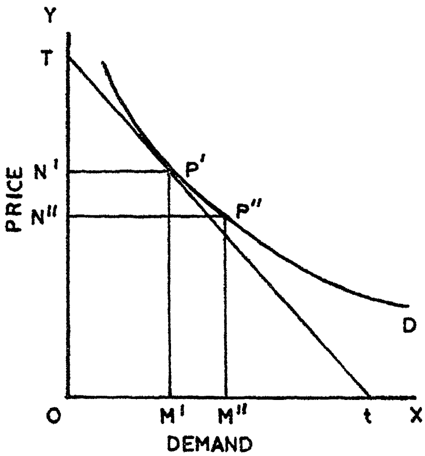


Diagram 2.4 (*a*)

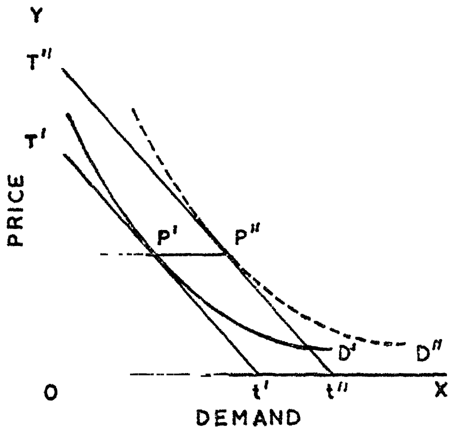


Diagram 2.4 (*b*)

In Diagram 2.4(*a*), *D* is the demand curve. The price is allowed to fall from ON' to ON'' leading to an increase of demand from OM' to

OM''. According to the measure of elasticity given above, elasticity of demand for the fall of price from ON' to ON'' is equal to $(N'N''/ON') \div (M'M''/OM')$. It can be shown as Marshall shows that it is equal to $P't/P'T$. If the fall of price is very small, as it is assumed to be for purposes of elasticity calculations, Tt would become a tangent to the demand curve at the point P'. Hence, we can say that the elasticity of a demand curve at any given point on it can be measured by drawing a tangent to it at that point and then dividing the lower part of it by the upper part.

In Diagram 2.4(*b*) two curves are drawn which a student of economics would call parallel. The elasticity of demand at price $P'M'$, for the curve D' is given by $P't'/P'T'$ while the elasticity of demand at the same price for the demand curve D'' is given by $P''t''/P''T''$. It will be obvious that the second measure of elasticity is smaller than the first. Thus curves that appear to be parallel have different elasticities for the same price.

If we take demand curves to be straight lines, as we often assume them to be for the sake of simplicity, we can lay down the relationships illustrated by Diagrams 2.5(*a*), (*b*) and (*c*) in regard to elasticity of demand.

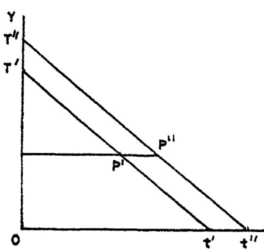


DIAGRAM 2.5 (*a*)

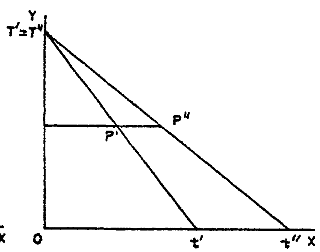


DIAGRAM 2.5 (*b*)

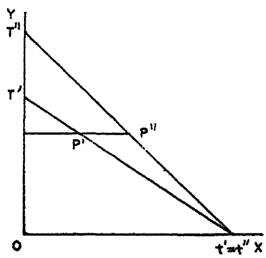


DIAGRAM 2.5 (*c*)

Of two parallel curves (straight lines can be called curves in this context), that which is farther away from the origin is less elastic than the other. This can be easily seen from Diagram 2.5(*a*) where the respective elasticities are $P''t''/P''T''$ and $P't'/P'T'$. Two curves that start from the same point on the Y-axis have the same elasticity at all corresponding prices. This can be seen in Diagram 2.5(*b*) where $P'P''$, which is parallel to the base $t't''$, divides the two sides of the triangle $T't't''$ in the same proportion. Of two straight-line curves that meet the X-axis in the same point, that which is farther away from the origin is less elastic. Diagram 2.5(*c*) shows this clearly. $P''t''/P''T''$ is smaller than $P't'/P'T'$.

It is essential for a student of modern economics to know all this about elasticity of demand. There is a strong temptation to judge elasti-

city of demand from the mere slope of curves. Dealing with the cobweb theorem, a student often makes the mistake of saying that when the demand curve (a straight-line curve) is more elastic than the supply curve, price oscillates but tends to converge to a stable level. One should, to be mathematically correct, say that when the slope of the demand curve is gentler than that of the supply curve, price converges to an equilibrium level.

Let us now diagrammatically show how elasticity of supply can be measured. In Diagram 2.6, S is the supply curve, positively inclined.

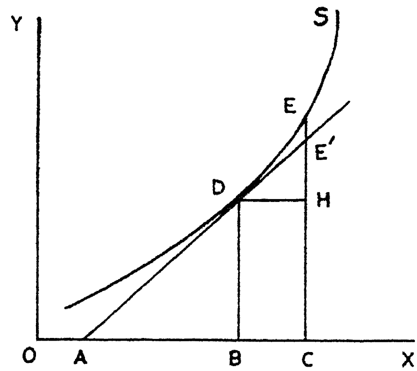


DIAGRAM 2.6

When the price rises from DB to EC supply increases from OB to OC. According to our formula, the elasticity of supply is given by $(BC/OB) \div (EH/HC)$ which is approximately equal to $(BC/OB) \div (E'H/HC)$. It can be shown that this is equal to AB/OB. When the points D and E are close together, i.e. when the price rises very slightly, the line DE tends to become tangential to the supply curve at the point D. We can, therefore, say that elasticity of supply is given by AB/OB, where OB is the supply at the price at which elasticity is calculated and AB is the intercept made on the X-axis by the tangent to the curve and the perpendicular dropped from the point at which the elasticity is calculated.

When the tangent passes through the origin O, AB and OB become equal so that elasticity of supply is 1. When the tangent cuts the X-axis to the left of the origin, AB becomes greater than OB so that elasticity of supply is greater than 1.

The elasticity of a curve, as we have seen, shows the response of one coordinate to a change in the other. There is a certain way in which we evaluate that response. We have taken the formula for measuring it from Marshall and we have also used his method of geometrically demonstrating it. So far the elasticity of a curve considered is what might be called point-elasticity. We now pass on to the arc-elasticity of

a curve. It was Dr. Dalton who introduced the concept of arc-elasticity of demand in economics.

ARC-ELASTICITY

Arc-elasticity refers to a finite change of price. When a price falls or rises in a market it does so by a finite amount, i.e., not by a mathematically infinitely small amount. Let us see how we can measure elasticity of demand when the price of commodity falls by a significant amount.

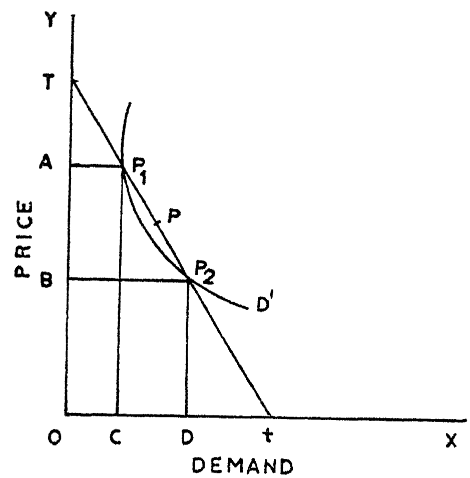


DIAGRAM 2.7

In Diagram 2.7 price falls from OA to OB and the demand increases from OC to OD. We draw a line to pass through the two corresponding points on the demand curve, P_1 and P_2. We have to find some measure for arc-elasticity of demand corresponding to the measure for point-elasticity of demand. We want to have a corresponding measure for the sake of harmony. Both point-elasticity and arc-elasticity are elasticities and so they must have common essential characteristics. The characteristics of point-elasticity are that (1) total expenditure on the commodity concerned remains unchanged when elasticity of demand is unity; (2) the measure of elasticity is the same whether we consider a rise or a fall of price, and (3) that it is independent of the units in which we measure price and demand. If arc-elasticity has to have these characteristics it should be equal to Pt/PT where P is the middle point of P_1P_2.

It will be seen that if the arc of the demand curve becomes smaller and ultimately reduces itself to a point, the point P would lie on the

demand curve and the arc-elasticity would convert itself into point elasticity. One who wants to read more about arc-elasticity should consult the works of Professor R.G.D. Allen, Professor A.P. Lerner and Dr. H. Dalton. We shall not give here the formula for elasticity of demand in terms of differential calculus as this is premature at this stage. Our object is not to study mathematical economics; for that many good books are available. We want to know something about the tools and techniques which are employed in modern economics.

KINKED CURVE

We have observed earlier that a curve can be continuous or discontinuous and smooth or crooked. One particular kind of crooked curve is what is called a kinked curve. In the study of imperfect competition, or what is more appropriately called impure competition, examples of kinked demand curves occur. Diagram 2.8(*a*) represents a kinked demand curve.

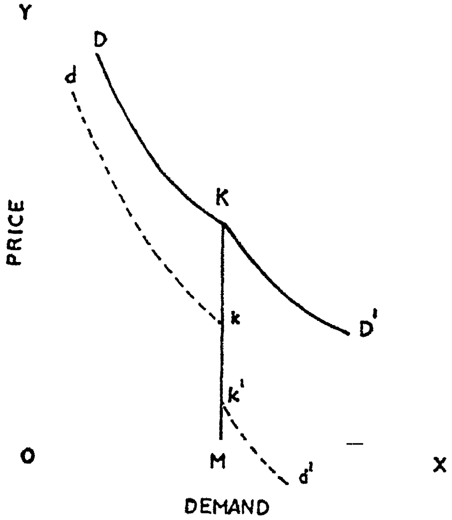


DIAGRAM 2.8 (*a*)

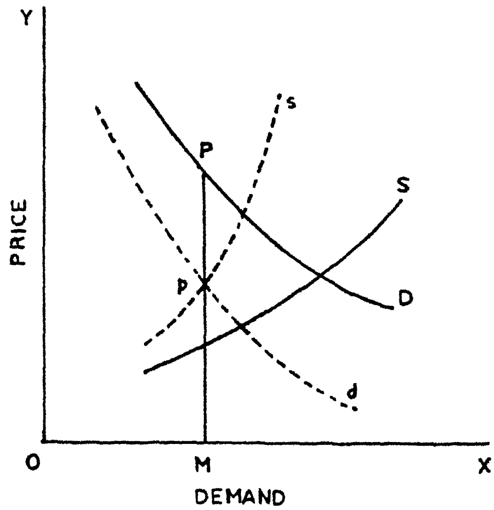


DIAGRAM 2.8 (*b*)

In Diagram 2.8(*a*) *DKD'* is the demand curve. It has a kink at the point *K* showing that at that point the fall of price with increased demand is precipitated. If we draw the corresponding marginal revenue curve it comes out to be *dk*, *k'd'*. It has, thus, two arms with a break in between. Marginal revenue is the addition made to total sale-proceeds by selling a little more of the commodity. More about this later.

A kinked demand curve is, it is maintained, likely to be met with in case of monopolistic competition. Suppose the price is *KM* and the demand *OM*. If now a seller were to reduce the price the demand would increase only slightly as this move of his is likely to be confronted with a similar move on the part of his rivals. But if he were to raise the price his rivals would not follow suit. The result would be a considerable reduction in the demand for his goods. Such a kinked demand curve, which is the result of the fact that price reduction is matched by similar reduction by others while a price rise is not likely to be so matched, results in a gap in the marginal revenue curve. The marginal revenue suddenly falls from k to k'.

In Diagram 2.8(*b*) there is a smooth demand curve *D* with its continuous marginal revenue curve *d*. We have also drawn a supply curve which is the average cost curve (*S*) and the corresponding marginal cost curve (*s*). As a student of economics knows, price is determined by the intersection of the marginal cost and marginal revenue curves. Hence price is *PM* and output is *OM*.

If the marginal cost curve were to pass in between k and k' in Diagram 2.8(*a*) there would be no real point of intersection of the marginal cost and marginal revenue curves. If then a price equal to *KM* were charged the seller's marginal revenue would be in excess of marginal cost. The seller would then be in a position to offer his goods at a slightly lower price, producing and selling the same amount as before. In ordinary circumstances such a change of price would be opposed by a similar cut in price by his rivals. But as this seller now does not sell a larger amount than before he would not cut into the sales of his rivals. There would be no matching reduction of the price charged by others.

The Indifference Curve

Another use of curves is to show a person's indifference to different consumption-goods (or combinations of them) and different production-goods (or combinations of them). A fillip to the use of curves for this purpose was given by the desire to understand the behaviour of economic units without assuming that utility was measurable. We shall, in its proper place, say something about measurability of utility; here we are concerned with the description of indifference curves.

Let us take consumption first. Suppose there are two commodities, *X* and *Y*, available to a consumer in various amounts. Let the commodity *X* be measured along the *X*-axis and the commodity *Y* along the *Y*-axis. In Diagram 2.9 we have drawn four curves. The point *A* on curve 2 represents On_1 quantity of *Y* and Om_1 quantity of *X*. This combination of the two commodities has some utility for the consumer. The curve 2

is so drawn as to give the different combinations of the two commodities yielding the same utility to the consumer. The point *B* representing On_2 of *Y* and Om_2 of *X* is on the same curve 2. The utility of this combination of goods is, therefore, the same as the utility yielded by the combination represented by the point *A*. As the consumer would be indifferent between these two combinations, curve 2 is called an indifference curve.

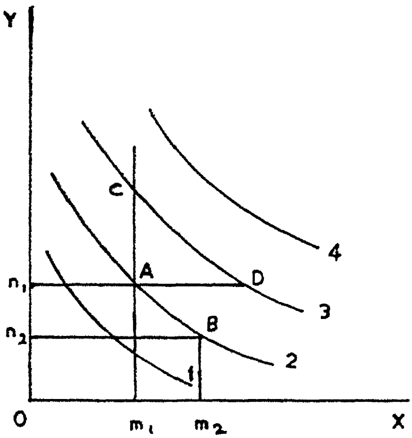


DIAGRAM 2.9

The point *C* shows a different combination of goods. The quantity of *X* is the same as before while that of *Y* is increased by the amount *AC*. This combination must, therefore, give more utility to the consumer. The curve 3 is again so drawn as to represent by various points on it different combinations yielding the same utility. The utility, it must therefore be obvious, goes on increasing as we pass on to higher and higher curves.

PROPERTIES OF INDIFFERENCE CURVES

In all normal cases, an indifference curve is negatively inclined and is convex to the origin. We have explained the meaning of the word convex as also that of negatively inclined. The negative inclination is due to the fact that, to keep utility constant when the quantity of one commodity is decreased, that of another has to be increased. This, again, is due to the fact that the possession or consumption of additional units of a commodity gives utility to the consumer no matter what amount he already has. In other words, we assume that the marginal utility of each commodity is positive.

The convexity of an indifference curve shows that, as the consump-

tion of one commodity is decreased, that of another has to be increased by greater and greater amounts. This will be so when the marginal utilities are decreasing or, to put in other words, when the marginal rate of substitution is increasing (according to Hicks, when the marginal rate of substitution is diminishing).

If, however, we assume that the utility of a commodity becomes negative after a time (or, as Hicks would say, if the marginal utility of one in terms of the other commodity becomes negative) an indifference curve would have to be drawn turning upwards after substitution has proceeded up to a certain limit. Such an indifference curve is shown in Diagram 2.10.

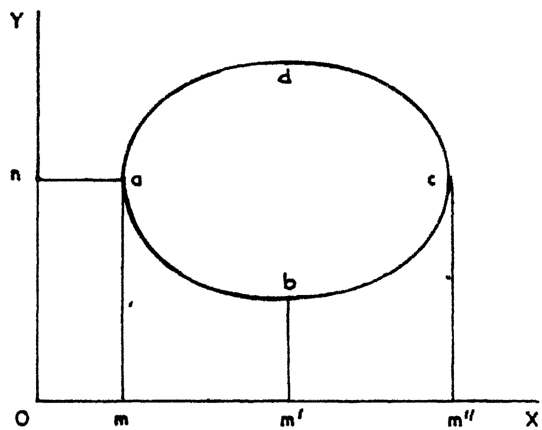


DIAGRAM 2.10

From the point *a* to the point *b* the curve runs as before; the marginal utility is positive and diminishing (in Hicks's terminology, the marginal rate of substitution of one commodity for the other is diminishing). After the point *b* is reached the marginal utility of *X* becomes negative. As the consumer increases the quantity of *X* he has to add to the consumption of *Y* to maintain total utility at the former level. The curve, therefore, begins to rise. Above the value *Om'* of *X* its marginal utility becomes negative. The curve continues to rise till the point *c* is reached. At this point the marginal utility of *Y* becomes negative. As the consumption of *Y* increases that of *X* has now to be decreased to prevent the total utility from decreasing. The loss due to the negative utility of *Y* has to be neutralised by decreasing the loss caused by the negative utility of *X*. From the point *c* onward *Y* increases and *X* decreases. The curve, therefore, begins to turn backward from the point *c* till the upper point *d* is reached. At this point a further decrease of *X* causes some loss of utility (instead of loss of disutility), to neutralise which the loss caused by the

consumption of Y has to be decreased. Y has, therefore, to be decreased after the point d is reached; the indifference curve begins to fall backward. This fall continues till the original point a is reached. Thus, when marginal utilities keep on diminishing till they eventually become negative an indifference curve becomes a closed figure.

As before, we can have a number of such indifference curves, each indicating a certain amount of utility. The farther away the convex and negatively-inclined part of the indifference curve from the origin the greater the utility it represents.

Indifference Surface

A curve has two dimensions whereas a surface has three. In the case of an indifference curve, one dimension measures the quantity of one commodity and the other dimension the quantity of the other commodity. In the case of a surface, there is the third dimension available to measure utility. In the case of an indifference curve the utility is not diagrammatically shown. It shows only that the farther away the curve from the origin the greater the utility. But now, in the case of an indifference surface, the third dimension can be utilized to show utility. This is done in Diagram 2.11.

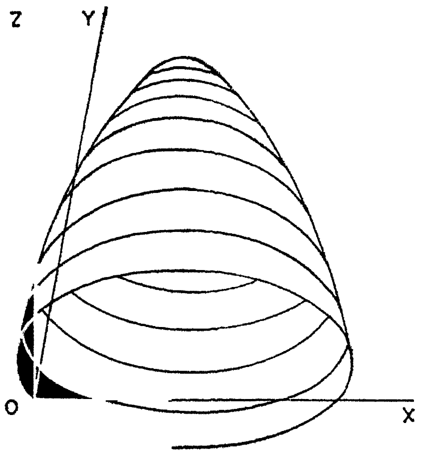


Diagram 2.11 (*a*)

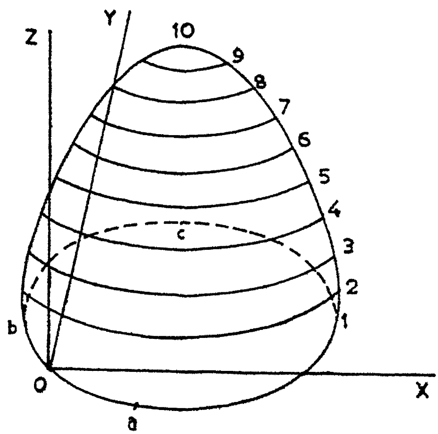


Diagram 2.11 (*b*)

In Diagram 2.11(*a*) the indifference surface is shown as it would appear when seen from within and the bottom. In Diagram 2.11(*b*) it is shown as it would appear when looked at from outside and from above. Diagram 2.11(*b*) is the more usual way of representing an indifference surface. Along the X-axis is measured the quantity of X and along the

Y-axis the quantity of Y. The utility represented by an indifference curve is measured along the Z-axis. The diagram thus becomes three-dimensional. Diagram 2.11(c) shows a more realistic outside view of the surface.

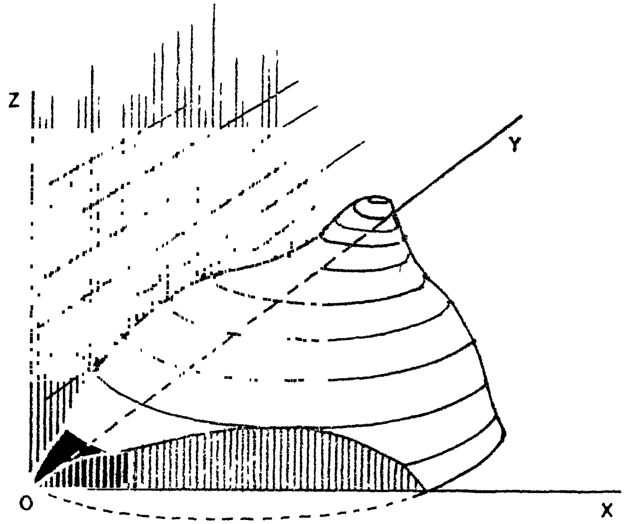


DIAGRAM 2.11 (c)

Diagram 2.11(b) shows ten indifference curves of which that marked 1 rests on the bottom: its Z-coordinate is zero. It shows quantities of X and Y which yield zero utility. The curve naturally passes through the origin. The curve 2 is a little above the base (the X-Y plane) and so it has Z-dimension also. The other indifference curves are higher and higher up the Z-scale and the highest indifference curve, 10 in the diagram, dwindles to a point and corresponds to the point M in Diagram 2.9. The quantities of X and Y represented by curve 10 in Diagram 2.11(b) are the quantities that yield the maximum utilities.

It will be seen that a part of the indifference surface is outside the X-Y quadrant, i.e., some of the indifference curves project to the left of the Y-axis and below the X-axis. Take, for instance, the curve 1. It passes through the origin at which point the quantity of X is zero and so also of Y. The utility is naturally zero. The point a represents a positive quantity of X and a negative quantity of Y. The utility of X is just balanced by the disutility of Y. (This exposition assumes that the consumer has a stock of X and Y with him). The point b represents a positive quantity of Y but a negative quantity of X. The utility of Y is just balanced by the disutility of X. Disutilities are suffered on account of the consumer losing, rather than getting, some amounts of the commodity concerned.

The point *c* on the indifference curve 1 represents positive quantities of both *X* and *Y* and yet the combination yields no utility. This is because the utility obtained from the positive quantity of *X* is just neutralised by the excessive quantity of *Y*. The total utility of *Y* is negative and not only its marginal utility.

One can draw diagrams of indifference surfaces in cases of constant and increasing marginal utilities but as they have no practical importance we shall not consider them here.

Production Indifference Curve

A consumer can be indifferent between combinations of goods while a producer can be indifferent between combinations of factors of production. A consumer is indifferent when the utility to be obtained is the same in all combinations. A producer is indifferent when the output to be obtained is the same in all combinations. In the case of consumption, we took two commodities into consideration as geometry allowed us to use only two axes (reserving the third for utility); in the case of production we can take two factors of production into consideration. Production requires all the factors of production and, therefore, when we take only two factors into account one of them must represent all the other factors. For example, one of the two factors can be labour and the other capital which can be taken to include organisation and enterprise also. Alternatively, one factor can stand for labour and organisation while the other can include capital and enterprise. Land is not a separate factor; it is merely an aspect of other factors.

In Diagram 2.12, let labour be measured along the *X*-axis and capital along the *Y*-axis. The indifference curve is again convex to the origin and negatively inclined. Its convexity shows that, as we decrease the amount of one factor, we have to increase the other factor by greater and greater amounts. Its negative inclination is due to the fact that, to keep output constant, a decrease of one factor has to be balanced by an increase of the other factor.

As in the case of consumption indifference curves, the middle curve in Diagram 2.12 yields more output than the lowest curve; or, to be more precise in our expression, the middle curve has been drawn to show combinations of factors that produce greater output than the combinations represented by the lowest curve. Similarly, the highest curve has been drawn to show combinations that yield still greater output.

Some consumption indifference curves were found to cut the axes; some of them did not. But a production indifference curve cannot cut either of the axes for the reason that it is not possible to produce anything with only one factor or, without employing all the factors. The

curves in the diagram above, therefore, tend to meet the axes at infinity: they are asymptotic to the axes. Whereas in consumption indifference curves (assuming the possibility of negative marginal utility), the highest curve shrinks to a point, in the case of production indifference curves this does not happen. Every curve extends itself to infinity at both ends. This however assumes that it is possible to substitute one factor for another although with greater and greater difficulty. This may not always be possible. After a certain minimum of a factor is reached, it is not possible further to reduce its quantity, substituting the other factor for it. In such cases we have to increase both the factors at some stage or other in order to keep the output unchanged. This is similar to the case of consumption indifference curves. For, here also a stage is reached

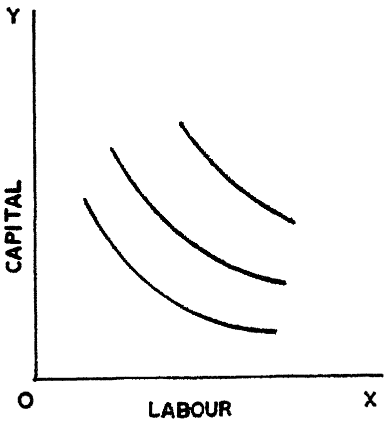


DIAGRAM 2.12

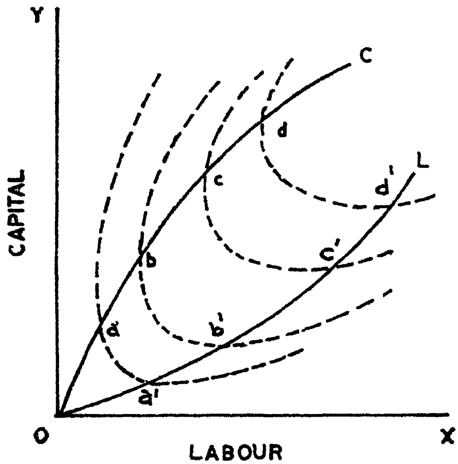


DIAGRAM 2.13

when the productivity of a factor (when combined with a certain—small—quantity of the other factor) becomes negative. There is, however, one important difference between production indifference curves and consumption indifference curves, namely, that the former are not closed figures. This feature is explained by the fact that whereas in the case of consumption both commodities can yield negative marginal satisfaction simultaneously within certain limits, in production both the factors cannot yield negative marginal productivity simultaneously. Diagram 2.13 shows the shape of an indifference curve when the productivity of a factor becomes negative.

The curves *L* and *C* in Diagram 2.13 indicate the points at which productivities of labour and capital, respectively, become negative. Between the limits prescribed by these curves, the production indif-

ference curves have the normal shape of consumption indifference curves when the marginal utilities are positive.

Use of Consumption Indifference Curves

The most important use that we can claim for indifference curves is that which arises from the doubtful nature of propositions about measurability of utility. If utility is not measurable, most of the theories explaining consumers' behaviour, it is believed, become of doubtful value. To give an example, Marshall's demand curve technique, it is generally believed, assumes that utility can be measured. It further assumes that the utility of money is constant. We shall have something to say in support of Marshall and in favour of the measurability of utility later on. Here let us merely observe that it is claimed that the indifference curve technique is not dependent on the assumption of measurable utility.

The basic problem of consumption-theory relates to the determination of quantities of goods that a consumer would buy at given prices. Marshall said that a consumer buys that amount of a commodity the marginal utility of which is equal to the price. This simplified way of explaining consumers' behaviour implies constant utility of money and also, in a way, measurability of utility. Hicks and Allen are responsible for the revival of the interest of economists in indifference curves. The behaviour of a consumer (really a buyer) is explained with the help of indifference curves as described below.

There are five indifference curves in Diagram 2.14. The commodity X

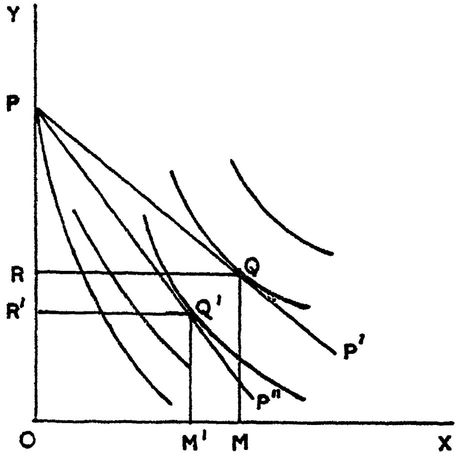


Diagram 2.14

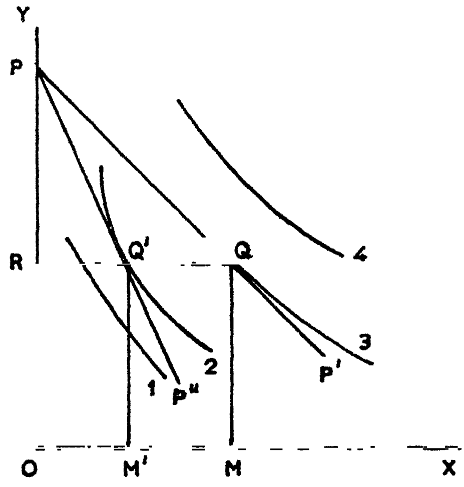


Diagram 2.15

is measured on the horizontal axis and the commodity *Y* on the vertical axis. If *Y* is money, the case becomes easily comparable to that of purchase and sale of *X*. Let us, however, speak in terms of *Y*. Suppose the consumer under consideration (one whose indifference curves are drawn in the diagram) has *OP* of *Y* to start with. Let the rate of exchange of *Y* for *X* be given by the slope of the line *PP'*. Now the data given, therefore, are (1) the utilities of different amounts of *X* and *Y*, indicated by the indifference curves, (2) the amount of purchasing power in the hand of the consumer, *PO* of *Y*, and (3) the price at which *X* can be bought, the slope of the line *PP'*. How much will the consumer buy and why? The consumer will buy *OM* of *X*, giving for it *PR* of *Y*—the slope of line shows that *OM* of *X* can be bought for *PR* of *Y*. He will buy this amount because he can then pass on to the highest possible indifference curve, i.e., he gets maximum utility.

So, given the data as enumerated above and making use of the principle that a consumer wants to maximise his utility, we can determine the amount that he will buy.

Now suppose the price of *X* in terms of *Y* rises. This can be indicated by the line *PP'* being given a steeper slope. Let the price be raised as indicated by the line *PP"*. The consumer will now buy *OM'* of *X* giving for it *PR'* of *Y*. He, therefore, buys a smaller amount of *X* and has to spend a larger amount of *Y*. He buys now *OM'* of *X* because by buying that amount he is able to maximise his utility, i.e., to pass on to the highest possible indifference curve.

Elasticity of Demand

We know that when the (point) elasticity of demand is unity the total expenditure of money on the commodity bought remains unchanged. In the above diagram, therefore, if the elasticity of demand is one, the total amount of *Y* spent in buying *X* should remain the same. In Diagram 2.14 if we go from *Q'* to *Q* the expenditure of *Y* decreases from *PR'* to *PR*. The elasticity of demand is, therefore, less than one. In Diagram 2·15, the rise of price from *PP'* to *PP"* keeps the expenditure of *Y* unaltered. The elasticity of demand is, therefore, equal to one. The rise of price results in the decrease of demand for *X* to *OM'* but the supply of *Y* has remained the same. It will be seen that the slope of indifference curve 3 at the point *Q* is the same as the slope of the line *PP'* while the slope of indifference curve 2 at the point *Q'* is the same as the slope of the new price line *PP"*. It can be shown that, when the elasticity of demand is greater than one, the expenditure of *Y* increases when the price falls. In that case *PP"* would be to the right of *PP'*, *PR'* would be greater than *PR*.

DEMAND CURVE AND INDIFFERENCE CURVE TECHNIQUES

A demand curve shows the amounts of a commodity that would be bought at different prices. In a demand curve diagram the price line is horizontal; in an indifference curve it is inclined (except when the price of one in terms of the other is either zero or infinity) at an angle to the X-axis that is not 90° or 180°.

There must be something in a demand curve to make it what it is. If it just says that so much would be bought at such and such a price it gives us what is required to be determined. Hence, the demand curve is identified with the marginal utility curve. In Diagram 2·16 the demand curve is, therefore, taken to be the marginal utility curve. At price *OP*, *OM* quantity of the commodity is bought or demanded because the utility is then maximum. And the utility is maximum because price and marginal utility are equal.

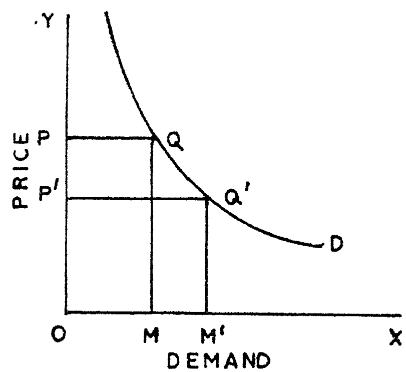


DIAGRAM 2.16

The point to be particularly noted here is that marginal utility is measured in terms of money (price-units) and since the measuring rod is ordinarily expected to be constant (of unchanging length), the marginal utility of money is assumed to be constant. This is considered to be an objectionable feature of the demand curve technique adopted by Marshall. From the point of view of theory it certainly is a point against this technique but, from the practical point of view, it is not. Money has a far more constant marginal utility than any specific commodity.

However, it is possible to make this technique theoretically fool-proof also. We can suppose that the marginal utilities represented by the demand curve are measured, not in terms of a constant supply of money, but in that of a varying supply of it. After the first unit of the commodity

is purchased there is a smaller amount of money left with the buyer. The second unit is, therefore, bought with money that has a higher marginal utility. Accordingly, adjustments are made in the amount that is shown as demanded. In simple words, the demand curve can be taken to incorporate the changing utility of money.

Let us now see how demand curves can be constructed out of indifference curves.

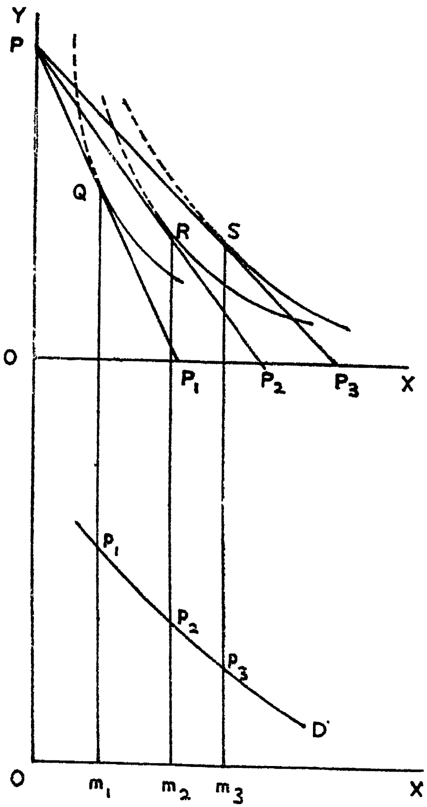


DIAGRAM 2.17

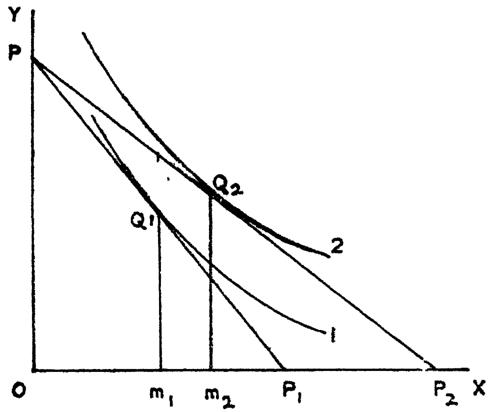


DIAGRAM 2.18

In Diagram 2.17 the upper figure shows the amounts of commodity X that would be purchased at prices shown by lines PP_1, PP_2 and PP_3. The quantities of X demanded are then marked off on the X-axis in the lower figure. The price given by the price line PP_1 is OP/OP_1. This is marked on the vertical line through m_1; p_1m_1 being equal to OP/OP_1. In the same way, p_2m_2 and p_3m_3 are also marked in the lower figure. If we join the p's we get the familiar demand curve.

The demand curve D here is one that does not assume constancy of marginal utility and, since it is derived from indifference curves, we

can say that the demand curve technique and the indifference curve technique yield the same result, provided care is taken to free the demand curve from the assumption of constant marginal utility of money.

In Diagram 2.18 two indifference curves are drawn with the vertical distance between them constant throughout. We can call them vertically parallel. Such curves indicate the fact that the marginal utility of the commodity *Y* is independent of the quantities of *Y* and *X* already in the possession of the consumer or already consumed by him. The more relevant point for our consideration is the fact that it is independent of the amount of *Y*. In technical language, this simply means that the marginal utility of *Y* is constant. This is precisely what Marshall assumed for purposes of his demand curve. When the price line changes from PP_1 to PP_2, i.e., when the price of *X* falls, demand for *X* increases just according to Marshall's demand curve.

The other point of difference between the demand curve and the indifference curve techniques, namely, that the former depends on measurability of utility while the latter does not, will be examined later when we take up the question of measurable utility.

Production Transformation Curve

A production transformation curve is also a variety of indifference curve. Here, the producer is indifferent to various combinations of products. His indifference will depend, among other things, on the relative prices of the products. Let us suppose our producer can produce two products *X* and *Y* with a given quantity of resources. The resources can be taken to be either money or factors of production. He has the option of utilising either the entire resources or a part of them. If he spends the resources

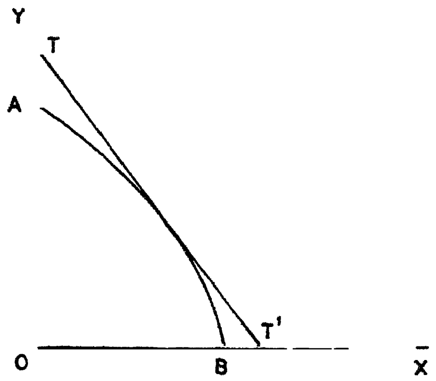


Diagram 2.19

(as fully as the technique of production will allow) on the production *Y* he can get *OA* of output (always per unit of time). If, on the other hand, he has to spend the whole amount on the production of *X* he would be able to turn out *OB* amount. He can utilise a part of the resources on the production of *X* and part on the production of *Y*. The manner in which the outputs of *X* and *Y* would vary in all ordinary cases is shown in Diagram 2.19.

The transformation curve *AB* is negatively inclined and concave to the origin, showing that as the output of one commodity increases that of another decreases and that it decreases by greater and greater amounts. The rate of transformation or substitution (of one commodity for another) is shown by the tangent to the curve.

Production Curve

Let us divide factors of production in two bundles. One bundle contains all but one factors of production while the other contains the remaining factor. Suppose the first bundle is fixed while the other is variable. To visualise this division, suppose that labour is variable while the other factors are kept constant. How will output vary, then, when labour is employed in varying amounts? The variation of output is shown in Diagram 2.20.

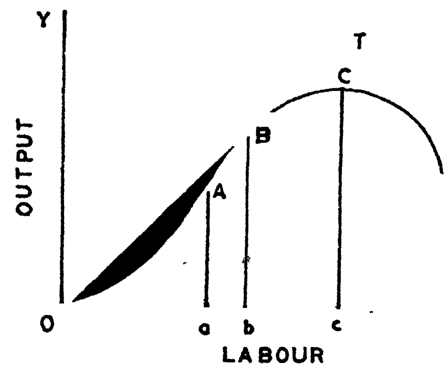


Diagram 2.20

As labour is increased, production increases (starting from output zero) at first at an increasing rate, then at a diminishing rate, till eventually it starts falling. The convexity downwards shows increasing rate of increase of output while the concavity of the curve shows decreasing rate of increase of output. The turning point *A* is called the point of inflexion.

The diagram shows that the marginal productivity of labour goes on increasing till *Oa* of labour is employed. Thereafter its marginal productivity begins to decrease. After *Oc* of employment the marginal productivity becomes negative.

We can also read from the diagram the average productivity of labour. For example, when *Oa* of labour is employed the average output is *Aa/Oa*. When *Ob* of labour is employed the average output is *Bb/Ob* and so on. The marginal output for any amount of labour employed is shown by the slope of the curve. At the point *B*, for instance, the slope of the curve is indicated by the tangent *T* passing through the origin. The marginal output when *Ob* of labour is employed is, therefore, *Bb/Ob*. It will be seen, therefore, that when employment is *Ob*, the marginal and average outputs are equal. If we draw average and marginal output curves they would intersect, therefore, when employment is *Ob*.

After *Oc* of employment marginal output becomes negative and, therefore, those employed in excess of that number cannot be called labourers. For, a labourer is a labourer by virtue of the fact that he is an agent of production and he is not an agent of production when he does not add anything to production.

Short-period and Long-period Curves

The average cost of production varies both in the short-period and the long-period. In the short-period it varies because of the variation of factors that can be increased or decreased in the short-period. In the long-period it varies because of the variation of factors that can be changed in the long-period.

In the short-period, it is not possible to change the technique of production. In the language of economics, organisation remains unchanged in the short-period. When, therefore, other factors are changed, the average cost of production changes. With one factor—organisation—fixed, the changes in other factors cannot be expected to keep average cost constant. Hence, in the short-period, average cost first decreases and then increases.

This change of average cost, when plotted, gives us a short-period cost curve. The term short-period has analytical, not historical or temporal, import. It is not really the passage of time that yields first decreasing and then increasing costs. The short-period cost curves show how the average cost would behave *if* production was to increase from zero to a finite amount; they do not show the behaviour of average cost over time.

In the long-period, average cost of production changes due to changes in all the factors of production, organisation included. But this change of organisation cannot be effected all at once. It requires time to change the

method of production. Hence, in the case of the long-period cost curve, there is a historical or temporal base. And yet it is possible to word our statements in terms of *if*'s rather than in those of *when*.

In Diagram 2.21 there are a number of short-period curves, i.e., curves in each of which organisation is kept constant and other factors changed. When the demand is low and only a small output is required, production is located on curve 1. When demand increases and a larger output is required, organisation is changed to suit a larger scale and production is located on curve 2. With every increase of demand, organisation undergoes a change and the short-period curve shifts to the right.

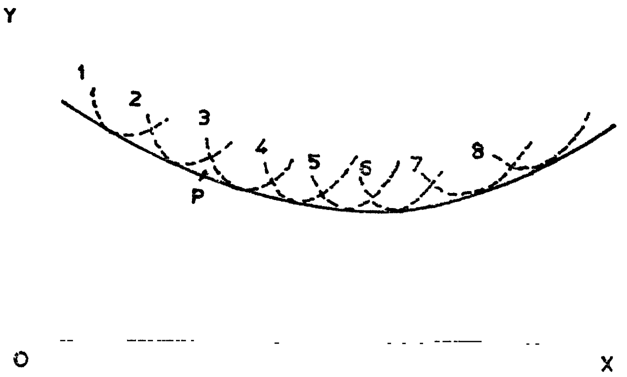


DIAGRAM 2.21

Out of these short-period curves we can construct a long-period curve by drawing an envelope. An envelope is a curve that envelopes all the short-period curves. Since the long-period curve is one in which even organisation changes, the envelope can, to that extent, stand for the long-period curve.

There is another method of constructing a long-period curve out of short-period curves and that is by drawing the locus of the lowest point of the short-period curve. If we join the lowest points of all the short-period curves (provided they are drawn very close to one another) we get a locus. Since, we may repeat, a long-period curve is one in which organisation also undergoes a change, the locus of the lowest point on the short-period curve can, in a way, be regarded as the long-period curve. The long-sweep plain curve in Diagram 2.22(*a*) is such a locus.

There is yet another way of constructing a long-period curve. We can consider the lowest portions of the short-period curves to form the long-period curve. Such a long-period curve is shown in Diagram 2.22(*b*) in thick lines.

Of the above three ways of drawing long-period curves, the third is the most realistic. Whether we take the analytical or the historical point

of view, production is always located on and proceeds along the third type of long-period curve. In the case of the envelope and the locus there are points on the curves that have no relevance to production. For instance the point *P* on the envelope and the locus is one where production is never located. The *Y*-axis of that point is never actually the average cost of production. Nor can one succeed in overcoming this difficulty by drawing the short-period curves very close to one another. For, organisation cannot be changed infinitesimally. If it could, organisation would lose the peculiarity from which arises the distinction between short-period and long-period curves.

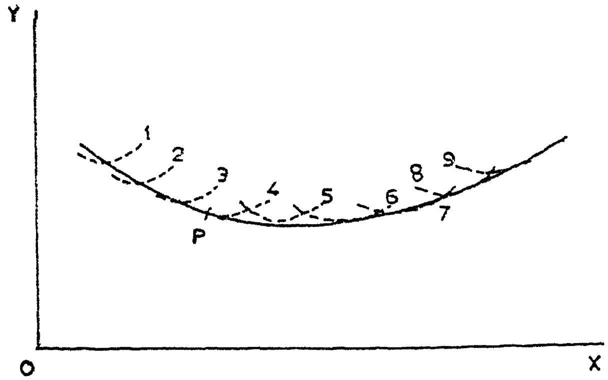


DIAGRAM 2.22(*a*)

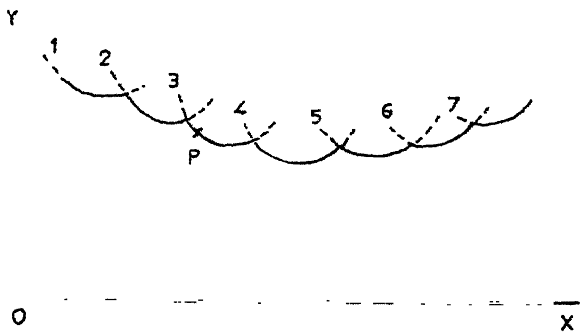


DIAGRAM 2.22(*b*)

We can construct from the third type of long-period curve of average cost a long-period marginal cost curve. This is done by drawing marginal cost curves relevant to the short-period average cost curves and then marking out the lowest portions as in the case of long-period average cost curves.

We have explained all the curves that a beginner would need to use while making a study of economics. We repeat that our object is not to study mathematical economics; all that we are attempting here is a study of simple tools of mathematics which the modern student of economics must know how to handle. In the next chapter, we shall see how curves can be represented by equations.

CHAPTER 3

Equations of Curves

What Does a Curve Represent?

A CURVE shows a relationship between two variables. The curves we have studied so far were drawn with reference to the axes of X and Y. Each point on such a curve measures certain distances from the X-axis and the Y-axis. These distances represent the magnitude of the variables whose relationship we want to study. Take, for example, the demand curve. The perpendicular distance of a point on the demand curve from the X-axis shows the price while its perpendicular distance from the Y-axis shows the demand at that price.

If the demand is 10 units at price 100 units we can express this relationship by saying that Y (price) is equal to $10X$ (demand). When we employ the letters X and Y to denote the magnitudes of the variables concerned we use the small letters rather than capitals. Thus we would say $y = 10x$. If the demand is always 1/10th of the price, $y = 10x$ would always be the relation between x and y, i.e., at all the points on the curve. We can then say that $y = 10x$ is the equation of the demand curve. But this cannot be a suitable equation for a real demand curve as it shows that demand is always one tenth of the price. Since, in reality, demand becomes a smaller and smaller proportion of price as price rises we have to have another equation which shows a varying relationship between price and demand.

In Diagram 3.1 there are two curves, both of which can serve as demand curves. The continuous curve shows that there is a maximum price OP at which there is no demand at all and that there is a maximum demand

OM at zero price. The dotted demand curve shows that, as in the other curve, there is a maximum price OP at which there is no demand, but that at zero price there is infinite demand. This curve meets the X-axis at infinity; we say in mathematics that the demand curve is asymptotic to the axis of X.

The curve PM has an equation of the form $y = c - bx + ax^2$. It will be seen from this equation that when x is zero y equals c and when y is zero x has a finite positive value. This equation, then, represents a normal demand curve. The curve is negatively inclined as the first derived function is negative, provided b is sufficiently large. The first derived function

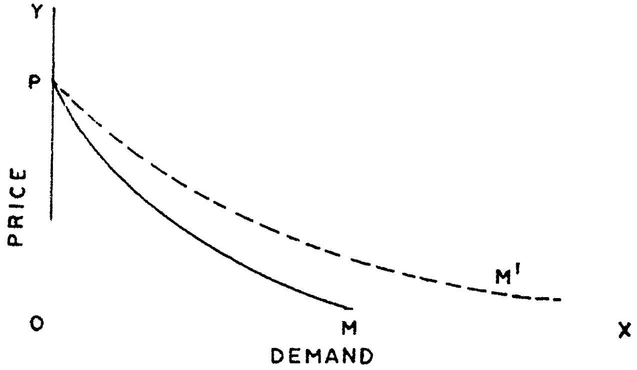


DIAGRAM 3.1

or, what is also called, the first differential of y is $-b + 2ax$. The meaning of *derived function* or *differential* and how it is worked out will be explained in another chapter. Further, the demand curve is convex to the origin and that is shown mathematically by the second derived function being positive. The second derived function or the second differential is $2a$.

We now take up the dotted curve which shows that there is no maximum limit to demand as it expands to infinity when price is zero. A suitable equation to this curve is $y = d/(g+x)$. It will be seen that when x is zero y is d/g, which in our diagram is equal to c. When y is zero x becomes infinity. This equation, therefore, traces out the dotted curve. Since this curve is also negatively inclined and is convex to the origin, its first differential must be negative and the second positive. And so they are; for, the first differential is $\dfrac{-d}{(g+x)^2}$ and the second, $\dfrac{2d}{(g+x)^3}$

COEFFICIENTS

A demand curve shows the relationship between price and demand. In this relationship, price is taken to be the independent variable. This means

that price varies first and demand varies as a response to it. This is why with lower price there is greater demand. Had demand been the independent variable, with greater demand the price would have been higher. Since there are two variables (dependent and independent), there are two quantities x and y in the equation. Besides these we have letters a, b, c, d and g. Of these, c is called a constant (it does not vary with x or y) and so are d and g. But b and a though constant, are called coefficients. In a way, they are as efficient as the variable x. Suppose $b = 10$ and $a = 1$. Then as x increases from 1 to 2, bx increases from 10 to 20, and ax^2 increases from 1 to 4. bx and ax^2 are accelerated or damped values of x and x^2. Price y is a function of x or demand. Or, to be more correct, x is a function of y as price is taken to be the independent variable. The function thus has three parts, one of which is constant while the other two are variables. With every change in x there is a change in y as far as the mathematical equation goes. But the second part of the function, namely bx, shows that the sensitiveness of y to changes in x is expressed by the coefficient b. If we take the third part of the function, ax^2, it shows that the sensitiveness of y to changes in x is expressed by the coefficient a multiplied by $2x$. This is because the first differential of ax^2 is $2ax$. We shall explain all this in another chapter.

Supply Curve

A supply curve shows the amounts that a seller or a producer of a commodity would be willing to offer at different prices. It shows, therefore, the cost of production, as price has to cover the cost if profit has to be maximised. In all cases, price tends to equal the marginal cost of production and, when competition is pure and perfect, also the average cost of production. We have, therefore, to find equation of average and marginal costs of production. From either of these we can find an equation for total cost of production. Or, we can reverse the process and begin with the equation of total cost of production and then derive from it those for marginal and average cost of production. We shall follow the latter method.

In Diagram 3.2 we have drawn three curves. The T.C. curve shows total cost, the A.C. curve shows average cost and the M.C. curve shows marginal cost. The total cost curve is always positively inclined because, with the increase of output, total cost increases. It is first concave and then convex downward showing that, at first, total cost increases at a decreasing rate and then at an increasing rate. The point of changeover, called the point of inflexion, is vertically above A on the X-axis. And at this point marginal cost is the lowest. This is because the marginal cost changes from decreasing to increasing costs at the point of inflexion.

The tangent to the total cost curve at the point C passes through the origin showing that the average cost (CB/OB) is equal to the marginal cost (CB/OB). We know that marginal cost at any point on the total cost curve is given by the slope of the curve at that point. And the slope is given by a tangent. Since at point C marginal and average costs are equal the A.C. and M.C. costs intersect vertically below the point C.

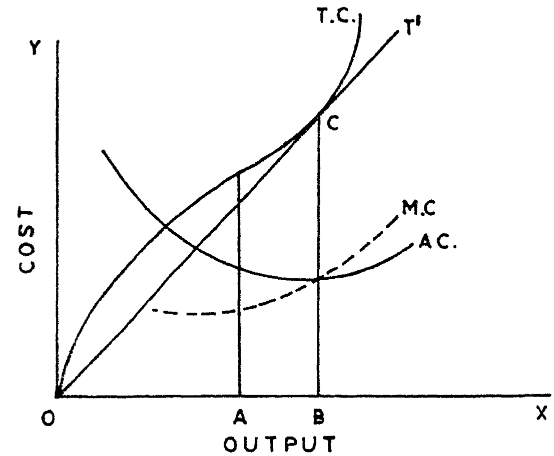


DIAGRAM 3.2

We can, therefore, say that for output OA marginal cost is lowest and the total cost stops increasing at a decreasing rate and begins to increase at an increasing rate. For output OB the average and marginal costs are equal, the average cost is lowest and the tangent to the total cost curve passes through the origin.

The most appropriate equation for a total cost curve of the above type is $y = Ax^3 + Bx^2 + Cx + D$. This curve must have the following properties. It must be positively inclined all through and it must be concave first and convex afterwards. To use the language of calculus (which will be explained in another chapter), the first differential of y must be positive and the second differential first negative and then positive. These conditions are satisfied when the coefficients A, C and D are positive and B is negative and has a comparatively large value. If we were to trace out such a curve we would get the T.C. curve of our Diagram 3.2.

The equation of an average cost curve can be derived from the above equation by merely dividing the expression by x, the output. We get then y/x or y(for the new curve) $= Ax^2 + Bx + C + D/x$. The equation for a marginal cost curve is given by the first differential to total cost. It is, therefore, dy/dx or y (for the new curve) $= 3Ax^2 + 2Bx + C$.

Sometimes, the total cost curve is drawn to show only increasing marginal and average costs. In that case it is throughout convex down-

ward. The equation of such a curve would be $y = Ax^2 + Bx + C$. Since it is positively inclined and its slope goes on increasing, all the coefficients (A, B and C) are positive. The equation is a quadratic one, i.e., it is a second degree function (x is squared). The equation of the total cost curve T.C. in Diagram 3.2 was a cubic one, a third degree function.

Discontinuous Curves

A demand curve is often discontinuous for the reason that demand does not increase every time the price falls. An infinite fall of price does not cause an increase of demand. Such a discontinuity can be called a jump-type of discontinuity. In economics we, however, ignore such discontinuities as they are difficult to handle with the tools of mathematics. It is easy to represent such cases diagramatically but not with equations. There is, however, another type of discontinuity for which we have simple mathematical expressions. This is when the jump-type of discontinuity can be located at finite points.

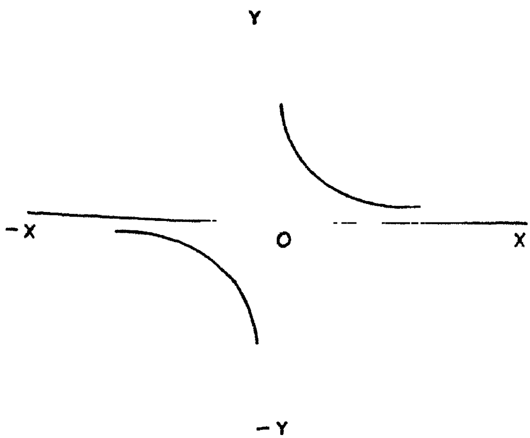


Diagram 3.3

There is another type of discontinuity which is found at infinity. Take for example the curve $xy = c$. This can be put in the form of $y = c/x$. When we trace out this curve we get the figure shown in Diagram 3.3. The curve has two branches, one in the first quadrant, the other in the third. As x decreases from a positive quantity to zero y increases to (positive) infinity. And as x approaches zero from a negative quantity y increases to (negative) infinity. We can say, therefore, that at x equal to zero y suddenly changes from one infinity to another. It becomes discontinuous for the value of x equal to zero but its discontinuity has points at infinity.

This curve, yielded by the equation $xy = c$, is called a rectangular hyperbola. Its significance in economics is that if a demand curve is a rectangular hyperbola the elasticity of demand is equal to 1 at all prices. But then in economics we are concerned only with that part of the curve which is located in the positive quadrant, i.e., where x and y both are positive.

INDIFFERENCE CURVES

Indifference curves show quantities of x and y yielding the same utility. As utility varies, the curves shift from the origin, approaching it as utility decreases and receding from it as utility increases. An indifference curve can become a closed figure if the marginal utilities of x and y (the commodities concerned) become negative after a certain amount is consumed. This was explained in the last chapter.

Let us take some examples of indifference curves and represent them by equations. It is convenient to take three types of curves, one represented by a parabola, another by a hyperbola and the third by a circle. This is only meant to illustrate how we can represent indifference curves by equations. The first two are not closed figures but a circle is.

An indifference curve that cuts the two axes and is convex to the origin can be represented by the equation

$$y = ax^2 + bx + c$$

This is an equation of a parabola. So that it may satisfy the requirements of an indifference curve a and c must be positive and b negative. The curve cuts the Y-axis at the height c and the X-axis at a distance

$$\frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

Actually the curve cuts the X-axis at two points but we limit the curve at the first point given by the above expression.

The vertex of this parabola is located in the fourth quadrant, i.e., its X-coordinate is positive and Y-coordinate negative. The usual equation of a parabola is $y^2 = ax$. But such a parabola has its axis horizontal and its vertex at the origin. When its axis becomes vertical (and for an indifference curve we want it to be vertical), y is changed into x and x into y so that the equation becomes $x^2 = ay$. And when the vertex is located in the fourth quadrant, i.e., when the origin shifts to left and upwards, we have to add a positive constant to y and a negative constant to x. With these changes, the equation becomes

$$(x - h)^2 = a(y + k)$$

which is of the from $y = ax^2 + bx + c$.

To draw a family of indifference curves we have to change the value of c. With a bigger value of c the curve starts from a higher point on the Y-axis. This is shown in Diagram 3.4. In Diagram 3.5 the indifference

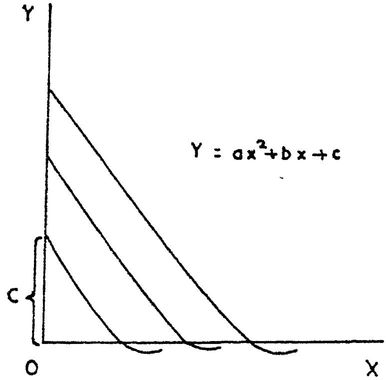


DIAGRAM 3.4

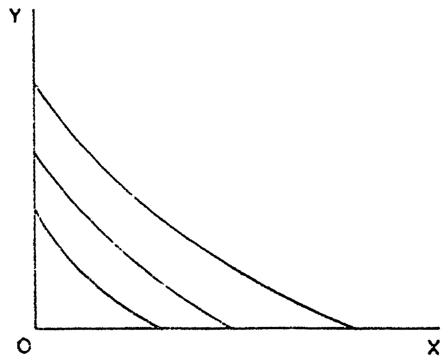


DIAGRAM 3.5

curves have a slightly different shape. To the eye they appear more stretched out as all hyperbolic curves are. We have to find appropriate equations for these curves. We have had an occasion to observe that $xy = a$, where a is a constant, is the equation of a rectangular hyperbola. Its economic significance, as we have already mentioned, is that such a curve represents a demand curve with unit elasticity of demand at all prices. Rectangular hyperbolas with the above equation are asymptotic to the axes, i.e., they do not cut them but tend to meet them at infinity. But in Diagram 3.5 they cut both the axes. We have, therefore, to change the equation slightly. In our diagram they can be shown to be asymptotic to the axes if they (the curves) are shifted to the right and upward. We have, therefore, to change the equation to $(x + b)\ (y + c) = a$, where b and c are constants showing the necessary shifts to the right and upward.

When marginal utilities become negative, the indifference curve becomes a closed figure which, for the sake of convenience, we can assume to be elliptical or circular. Let us assume that it is circular. The equation of a circle, that has the origin as its centre and radius a, is $x^2 + y^2 = a^2$. This is derived from the fact that the distance of the circumference from the origin (centre) is constant. When the indifference curve is a circle, its centre is not the origin; it is the point whose coordinates measure the amounts of the commodities whose marginal utilities are zero. The family of indifference curves consists of concentric circles (circles that have the same centre). As the centre of the circle is to the right and above the origin, we have to shift the origin to the left and downward. The above equation is, therefore, changed to

$$(x - b)^2 + (y - c)^2 = a^2.$$

Here b and c are constants which show the distances shifted. Tthe radius is a ; so by giving different values to it, we get a family of indifference curves.

Constant-Product Curves and Ridge Lines

In Diagram 2.13 we have drawn a family of constant-product curves which we have called production indifference curves. A suitable equation for a member of such a family of curves is $2Hxy - Ax^2 - By^2 = Z$. Here Z shows the constant output in each case and x and y are quantities of the two factors. It can be shown that this curve has a negative slope (is negatively inclined) within the ridge lines OC and OL but outside these lines the slope is positive because of the negative marginal productivity of factors as already explained in the last chapter. According to differential calculus, the slope is given by the first differential of y with respect to x which, in the above equation, can be shown to change its sign outside the values of x and y given by the ridge lines.

The equation of the ridge lines can be found by differentiating the constant Z with respect to x and y respectively and making the differentials equal to zero. We thus find the equations of the ridge lines to be $y = \frac{H}{B} x$ and $y = \frac{A}{H} x$ respectively. It will be seen that the ridge lines OC and OL are the loci of the turning points of the indifference curves. In general, the ridge lines need not be straight lines, but may be curves as in Diagram 2.13. In the case of the equation mentioned above, they are straight lines.

Symbols to Represent Equations

An equation that represents a curve, drawn with reference to the axes of X and Y, shows a relationship between two variables. Among the variables commonly featuring in economics are demand, supply, price, utility, costs and output. A curve can show the relationship between any two of them. The relationship is naturally quantitative and not qualitative. As a matter of fact it is doubtful if there can be anything like qualitative relationship, or at any rate it is arguable that in the study of a science it is only quantitative relationships that matter.

When two variables, represented by x and y, are related in a way that can be depicted by an equation, the relationship is called a functional relationship and each variable is said to be a function of the other. For example, in the equation $y = 2x^2$, y can be said to be the function of x.

And we can rewrite the equation in the form $x = \sqrt{\frac{y}{2}}$ and say that x is a function of y. Technically speaking, in such cases, x and y belong to an implicit function. Instead of numbers we can use letters and express a functional relationship by an equation such as $y = ax^2$. A more complicated relationship, as we have seen, can also exist between two variables. We can thus have an equation of the third, the fourth and higher degrees.

Instead of writing out the equation in arithmetical or algebraic form we can give a symbolic representation to it. We can say, for instance, that $y = f(x)$, which is read as "y is a function of x". When the relationship is so expressed one does not know what precisely is the nature of the relationship. But we can perform mathematical operations on such equations, expressing at each step the result by means of symbols. Ultimately, of course, we have often to express in numerical or algebraic form the conclusions reached.

Such symbolic representation of equations is useful particularly when we have to subject them to processes in the mathematics of calculus. What that is we shall see later.

A symbolic representation of an equation tells us only one thing, namely, that one variable is a function of another. But when the equation is not that of a curve it can express a relationship between many variables. Thus, we can have an equation such as $y = f(x,z,a,b\ldots)$ which tells us that there is a quantitative relationship between y and the other variables within the brackets. What type of relationship exists between the variables cannot be known merely from the symbolically represented equation.

When the equation is fully expressed in arithmetical form such as $y = 5x^2 + 3z$, it gives us detailed information about the relationships involved. It tells that y varies as five times the square of x and 3 times z. And yet a symbolically expressed equation has advantages. Its main advantage consists in the ease with which we can handle it and subject it to mathematical operations. When a symbolic equation is subjected to some mathematical operation the relationship between the variables or the consequences of such a relationship is preserved. But the fact remains that such an equation gives us very little information. We do not know the coefficients of the variables nor their power. But there is a way of conveying the information that one function or equation is different from another. That can be done by writing one equation as $y = f(x)$ and others as $y = F(x)$, $y = \phi(x)$, $y = \Psi(x)$, etc. Or, alternatively, we can suffix the 'f' with numbers such as f_1, f_2, f_3, etc.

CHAPTER 4

The Differential Calculus

What is the Differential Calculus?

IN THE LAST chapter we had occasion to observe that the equation of a curve can be expressed as $y = f(x)$ where x and y are two variables functionally related and that, when there are more variables, we can write the function as $y = f(a, b, c\ldots)$ though no curve can be traced out if there are four or more variables. We also observed that this symbolic representation of an equation is very convenient when we have to subject our equations to operations in the differential calculus. The time has now come to see what this calculus is.

When we say x and y are functionally related, it means that a change in one causes a change in the other. As scientists, we are interested in knowing the extent to which a change in one variable causes a change in the other. The differential calculus is a method of determining the rate at which one variable changes as a result of a change in the other variable. For example, in the equation $y = f(x)$, if x changes by one unit and y, in consequence, changes by two units, we can say that the rate of change is 2. This rate may be constant or itself changing—changing as the value of x changes. Such a calculus is of no service when there is no functional relationship between two variables, nor has it any use when the variables are not continuously variable. A real functional relationship can be said to exist for our present purpose only when every small change in x is conceivable and a corresponding change in y is also conceivable. The rules and methods by which we calculate the rate at which a function

(y, for instance) changes in consequence of a change in a variable (x, for instance) constitute the differential calculus.

It is important to note the meaning of the word *rate*. A rate is a function of two things. The absolute amount of change divided by the number of units, from which the change took place, is called the rate of change. But in the differential calculus, the rate of change of y due to change in x implies the above rate when the smallest possible change occurs in x. More about this later.

Importance of the Differential Calculus

In the study of every science we are concerned with magnitudes of entities we deal in. Further, there can be no knowledge worthy of the name science unless it relates one magnitude to another. The ultimate object of a science is to discover causal relationships between phenomena and for such a discovery it is necessary to be able to measure our entities quantitatively and find relationships between them. It is here that calculus assumes its importance as a necessary tool for the discovery of causal relationships between phenomena, which is the same thing as the establishment of laws and principles.

In economics, for example, we want to discover causal relationships between demand and price, between price and supply, between output and cost, between consumption and utility and so on. These relationships, as we observed earlier, must necessarily involve quantitative measurement of entities. It is of very little use for us to know that demand increases when price falls: we must know to what extent it increases when the price falls by one unit. It is in such cases and for such purposes that the method called the differential calculus has to be employed. The more exact and the more correct we want to be in our science, the more necessary it becomes to be able to relate one change to another and, therefore, to devise ways of calculating the rate at which a thing changes due to changes in another thing (the cause).

The Rate of Change Explained

Suppose we have a functional relationship between two variables x and y expressed by the equation $y = x^2$. If x is equal to 1, y also becomes equal to 1. But when x changes to 2, y changes to 4. It will be noticed that when x becomes double y becomes fourfold. Now let us calculate the rate of change in y. Suppose x increases not by one unit as in the previous example but by a quantity h (we use this symbol to designate a small change in x for purposes of calculating the rate of change). Then from x^2, y increases

to $(x+h)^2$. The absolute increase is equal to $(x+h)^2 - x^2$ and the rate of increase, i.e., per unit increase (a very small increase) can be expressed by

$$\frac{(x+h)^2 - x^2}{h} = \frac{2hx + h^2}{h} = 2x + h$$

If now h becomes very small and ultimately zero, this rate becomes equal to $2x$. The limiting value of the rate, i.e., when h becomes zero is given by $2x$.

We can proceed in the same way to find the rate of change in any other function. Suppose the function is $y = ax^2 + bx$. If x changes to $x + h$ the value of y becomes $a(x+h)^2 + b(x+h)$. And we get the rate of change equal to

$$\frac{a(x+h)^2 + b(x+h) - ax^2 - bx}{h} = 2ax + ah + b$$

And when h becomes equal to zero, this rate reduces to $2ax + b$.

Now it will be obvious that if h is zero, it means that x has not changed at all and when it has not changed there can be no question of rate of change of the function (i.e., of y). Hence this rate of change is called the limiting value of the actual rate of change of y, i.e., the rate of change approaches more and more closely to this value as h approaches the value zero. Hence for a given value of x (as x) the value of y tends to change at the above rate. This is the normal rate of change, indicating the tendency inherent in the value of y which expresses itself in some finite measure when x actually changes (which change naturally must be finite itself). This rate is called the derivative of a function. A function, then, has a derivative when it is a continuous function, i.e., a function of variables which are capable of continuous variations. Further, it is necessary that the rate should tend to a definite limit. In the above example, this definite limit is given by $2ax + b$.

Names and Notation

The rate of change of a function, or the instantaneous rate as it is called, or the limiting value of the rate that is found by the use of differential calculus, can be a function of the changing variable. In the above example the rate of change of y is $2ax + b$, i.e., it is a function of x (it varies with x). The limiting value of the rate of change of y is called the derivative. But it can also be called a derived function. And such a derived function or derivative can be represented by a notation. There are a number of ways in which the derivate of y is symbolically represented. They are:

$$\frac{dy}{dx}, \frac{d}{dx}f(x), f_x \text{ and } f'(x)$$

All these notations mean the same thing: all of them stand for the rate of increase of the function as x increases. When the function is known in terms of numericals or other symbols such as letters of the alphabet we can use the second of the above three ways of representing a derivative. For instance, if the function is $ax^2 + bx + c$ we can represent the derivative as follows:

$$\frac{d}{dx}\,(ax^2 + bx + c)$$

When the exact form of a function is not known we can use the last notation. Here we can distinguish one function from another by using different letters. Thus, one function can be written as $f(x)$, a second as $\phi(x)$, a third as $\theta(x)$ and the fourth as $\Psi(x)$.

Some Derivatives to Remember

It is useful and helpful to remember the derivatives of some common functions. From the meaning of a derivative and the way it is calculated it becomes obvious that the derivative of x^3 is $3x^2$ and of x^2, $2x$ and so on. We can generalise and say that the derivative of x^n is $n.x^{n-1}$. It can be shown in the same way that the derivative

$$\text{of} \quad \sqrt{x} = \frac{1}{2}\cdot\frac{1}{\sqrt{x}} \qquad (1)$$

$$\text{of} \quad x^{\frac{1}{n}} = \frac{1}{n}\cdot\frac{1}{x^{\left(\frac{1}{n}-1\right)}} \qquad (2)$$

$$\text{of} \quad \frac{ax+1}{x-1} = -\frac{a+1}{(x-1)^2} \qquad (3)$$

$$\text{of} \quad \log_e x = \frac{1}{x} \qquad (4)$$

$$\text{of} \quad x = 1 \qquad (5)$$

$$\text{of a constant} = 0 \qquad (6)$$

$$\text{of} \quad a^x = a^x \log_e a \qquad (7)$$

This will be proved later.

Let us remember that a derivative of a function shows the rate at which the value of the function increases when a variable increases. Thus when x increases the square root of x increases at the rate of one-half of the reciprocal of the square root of x. Similarly, the rate at which x increases when it does increase is one. And since a constant does not increase its rate of increase (its derivative) is zero. The rate at which log x increases with the increase of x is the inverse of x, $\left(\frac{1}{x}\right)$—we cannot understand this unless we know what logarithms are.

Logarithms

If $a^y = x$, then y is said to be the logarithm of x to the base a. This means that if a is multiplied by itself y times (i.e., there are y number of a's multiplied together), then we get x as the product. Hence, a is said to be the base, and y is the logarithm of x. A logarithm is, therefore, the power to which a given base is raised to give a certain value. It is useful to remember the following logarithms:

$$\text{if } a^y = x, \ \log_a x = y$$

$$\text{Since} \quad a^1 = a, \log_a a = 1 \tag{1}$$

$$\text{Since} \quad a^0 = 1, \ \log_a 1 = 0 \tag{2}$$

$$\text{If } a^{y_1} = x_1 \text{ and } a^{y_2} = x_2$$

$$\text{then} \quad y_1 = \log_a x_1 \text{ and } y_2 = \log_a x_2$$

$$\text{Also } x_1 . x_2 = a^{(y_1+y_2)}$$

$$\therefore \quad \log_a (x_1 \cdot x_2) = y_1 + y_2 = \log_a x_1 + \log_a x_2 \tag{3}$$

Similarly it can be shown that

$$\log_a \left(\frac{x_1}{x_2}\right) = \log_a x_1 - \log_a x_2 \tag{4}$$

$$\text{If} \quad x = a^y \text{ or } \log_a x = y$$

$$\text{then} \quad x^n = a^{ny} \quad \text{or} \quad \log_a x^n = ny = n \log_a x \tag{5}$$

Derivatives of Logarithms

Having seen what logarithms are, let us now explain how we can find the derivative or a derived function of a logarithm. We said above that the derivative of log x is $\frac{1}{x}\cdot$ We can work out the derivative by applying the usual process in the following way:

The derivative is

$$\frac{\log (x+h) - \log x}{h} = \frac{1}{h} \log \left(\frac{x+h}{x}\right) \qquad \text{from (4) above}$$

$$= \frac{1}{h} \cdot \frac{1}{x/h} \log \left(1 + \frac{h}{x}\right)^{x/h} \qquad \text{from (5) above}$$

$$= \frac{1}{x} \cdot \log \left(1 + \frac{1}{n}\right)^n \text{ where } n = \frac{x}{h}$$

In the limit, when h tends to zero, n tends to infinity. But it is known that when n tends to infinity, $\left(1 + \frac{1}{n}\right)^n = 2{\cdot}7182..,$ usually denoted by e.

Hence $\frac{1}{x} \log_e \left(1 + \frac{1}{n}\right)^n = \frac{1}{x} \log_e e$

$= \frac{1}{x}$ from (1) above

The above is the derivative of $\log_e x$, i.e., if y is the logarithm then $e^y = x$. Now let us find out the derivative of $\log_a x$. Now the base is a so that, if the logarithm is y, then $a^y = x$. There is a rule for a change of base which the student may remember. It is that

$$\log_a x = \log_e x.\ \log_a e.$$

Hence, to find the derivative of $\log_a x$, we must differentiate $\log_e x.\ \log_a e$.

$$\frac{d}{dx} \{\log_e x.\ \log_a e\} = \frac{d}{dx} \log_e x.\ \log_a e.$$

$$= \frac{1}{x} \cdot \log_a e$$

Now suppose $y = f(x)$.

Then $\log y = \log f(x)$ and $\frac{d}{dx} \log y = \frac{d}{dx} \log f(x)$

which is equal to $\frac{d}{df(x)} \log f(x) \times \frac{d}{dx} f(x)$

$$= \frac{1}{f(x)} \cdot \frac{f'(x)}{1}$$

We can now give the proof of the derivative of a^x, which was stated earlier to be $a^x.\ \log_e a$.

Put $\log_e a$ equal to y. Then $e^y = a$ or $e^{\log_e a} = a$.

Hence $a^x = (e^{\log_e a})^x = e^{x \log_e a}$

and $\frac{d}{dx}(a^x) = \frac{d}{dx}(e^{x \log_e a})$

$$= \frac{d}{d(x \log_e a)} (e^{x \log_e a}) \times \frac{d}{dx} (x \log_e a)$$

$$= e^{x \log_e a} \times \log_e a$$

$$= a^x \log_e a$$

Here, it will be seen, we have made use of the formula that

$$\left\{\frac{d}{dx \log_e a} (e^{x \log_e a}) = e^{x \log_e a}\right\}, \frac{d}{dx} (e^x) = e^x$$

The proof is as follows:

Put e^x equal to y. Then $\log_e y = x$.

Differentiating with respect to y, we get $\frac{1}{y} = \frac{dx}{dy}$

or $$\frac{dy}{dx} = y$$

or $$\frac{d}{dx}(e^x) = e^x$$

DERIVATIVES OF SUMS, PRODUCTS AND QUOTIENTS

Just as we can differentiate a function, we can differentiate a sum of two or more functions or their product. We can also differentiate the quotient of two functions.

Suppose $x = y + z$. Then the rule is that the derivative of $y + z$ with respect to x is given by the sum of the derivative of y and the derivative of x. Thus, $\frac{d}{dx}(y + z) = \frac{dy}{dx} + \frac{dz}{dx}$

The derivative of the product of y and z is given by y times the derivative of z plus z times the derivative of y. Thus,

$$\frac{d}{dx}(y.z) = y\frac{dz}{dx} + z\frac{dy}{dx}$$

The derivative of the quotient of y divided by z is given by z times the derivative of y minus y times the derivative of z and the result divided by z^2. Thus, $\frac{d}{dx}\left(\frac{y}{z}\right) = \frac{z\frac{dy}{dx} - y\frac{dz}{dx}}{z^2}$

These formulae can be proved by working out the rate of change as in the case of a simple function of x.

THE DERIVATIVE OF A FUNCTION OF A FUNCTION

If in the equation $y = f(x)$, x is itself a function of z such as, for example, $\phi(z)$ we can say that $y = f\{\phi(z)\}$

The derivative of y with respect to z, in such a case, is given by the derivative of y with respect to x multiplied by the derivative of x with respect to z. Thus, $\frac{dy}{dz} = \frac{dy}{dx} \cdot \frac{dx}{dz} = f'(x) \cdot \phi'(z)$

In the same way if $y = f(x)$, $x = \phi(z)$ and $z = \Psi(u)$

$$\frac{dy}{du} = \frac{dy}{dx} \cdot \frac{dx}{dz} \cdot \frac{dz}{du}$$

In an example let $y = (ax^2 + bx + c)^{\frac{1}{2}} = (Z)^{\frac{1}{2}}$

$$\frac{dy}{dx} = \frac{dy}{dz} \cdot \frac{dz}{dx}$$

$$= \tfrac{1}{2}(Z)^{-\frac{1}{2}} \cdot \frac{dz}{dx}$$

$$= \tfrac{1}{2}\,\frac{1}{\sqrt[2]{z}} \cdot \frac{2ax + b}{1}$$

$$= \tfrac{1}{2}\,\frac{2ax + b}{\sqrt[2]{ax^2 + bx + c}}$$

Higher Order Derivatives

We have explained how the derivative of a function is obtained. This is called the first order derivative or, if it is a function, we can call it the first derived function. Thus, the derivative of $5x^2$ is, as already explained, $10x$. This is called the first (-order) derivative of $5x^2$. Similarly, we can find the derivative of a cubic or higher order expression in x. These derivatives are themselves functions of x, which means that they contain the variable x. We call them the first derived function of x.

After performing such an operation on a function of a variable, we can repeat it. Thus, we can differentiate $10x^3$ and get $30x^2$. This is the first derived function of x. Now, we can differentiate $30x^2$ and get $60x$. This, then, can be called the second derived function. To use the most correct expression we would say that $60x$ is the second (-order) differential of the original function of x, namely, $10x^3$. One can go on repeating the operation and getting thereby derivatives of higher and higher order. In the above example the third differential of $10x^3$ is 60. As this is a constant the fourth differential would be zero.

The second derived function is represented by notations similar to those for first derived function. If $y = f(x)$ is a function of x, the second derived function can be represented by the notations,

$$\frac{d^2y}{dx^2},\ \frac{d^2}{dx^2}f(x) \text{ and } f''(x)$$

In the same way the third derived function can be represented by the notations,

$$\frac{d^3y}{dx^3},\ \frac{d^3}{dx^3}f(x) \text{ and } f'''(x)$$

The Meaning of Higher-order Derivatives

The first derivative shows the rate at which the dependent variable varies with the independent variable. Or, to express it in more general terms (applicable to implicit and explicit functions), the first derivative

shows the rate at which a variable varies when the other variable is changed. In the same way, the second derivative of a function shows the rate at which the first derivative changes.

Take, for example, the case of a functional relationship between the quantity of a commodity demanded and its price. Suppose the relationship is given by the function $p = 10x^2 + 20x$. (p: price; x: demand). The first derivative of this function is $20x + 20$. The price of the commodity, therefore, increases at the rate $20x + 20$. The second derived function is 20. The rate at which the first derivative varies is therefore, 20, meaning that the rate at which the rate of increase of price varies is 20.

The importance of higher order derivatives will become obvious as we proceed further. For instance, the maximum value of a function is given by the condition that the first derived function is zero and the second derived function is negative. Similarly, the minimum value of a function is given by the condition that the first derivative is zero and the second derivative is positive. These applications are explained below. Further, in the case of the point of inflexion, the point at which a curve changes its concavity into convexity or vice versa, the second derivative changes its sign, that is, it is zero at the point of inflexion.

Maxima and Minima of a Function

In Diagram 3.2 in Chapter III there are three curves, namely, the total cost curve, the average cost curve and the marginal cost curve. The marginal cost curve M.C. is shown passing through the lowest point on the average cost curve, A.C. The equation of the total cost curve there is T.C. $= Ax^3 - Bx^2 + Cx + D$. The equation of the average cost curve is A.C. $= Ax^2 - Bx + C + D/x$. At its lowest point the slope of the curve turns from negative to positive. But the slope is given by the first derivative of A.C., i.e., it is equal to $2Ax - B - D/x^2$. If this turns from a negative quantity to a positive quantity, at the turning point it must be zero. Hence at the lowest point on A.C. the first derivative is zero. Since the first derivative changes from negative to positive its value is algebraically increasing. Hence, the derivative of the first derivative must be positive. And this means that the second order derivative must be positive.

We thus find as conditions for a minimum of a function that the first derivative should be zero and the second positive. It can likewise be shown that for a maximum of a function the first derivative should be zero and the second derivative negative.

Slope of a Curve and the Tangent

In Diagram 3.2 in Chapter III we have drawn a line OT' through the origin which is tangential to the total cost curve at the point C. We know that the slope of a curve at a given point is indicated by the tangent to the curve at that point. Thus, at the point C the total cost curve has a slope indicated by the line OT'. But the slope of the curve at the point C shows the rate at which the Y-coordinate of the curve at C is increasing. It is the limiting value of the ratio of increment of y to the increment of x at the point C. And, as we know, that is given by the first derivative of the function, which in this case is $Ax^3 - Bx^2 + Cx + D$.

Hence, we can say that the slope of a curve at a given point is indicated by that of the tangent to the curve at that point (i.e., it is equal to the trigonometrical tangent of the angle made by the geometrical tangent with the X-axis) and also by the first derivative of the function.

Positive and Negative Slopes

A curve is said to have a positive slope when the geometrical tangent is positively inclined. And a geometrical tangent is said to be positively inclined when the trigonometrical tangent of the angle made by the geometrical tangent with X-axis is positive. In Diagram 4.1 there are two curves. The curve A is positively inclined while the curve B is negatively inclined.

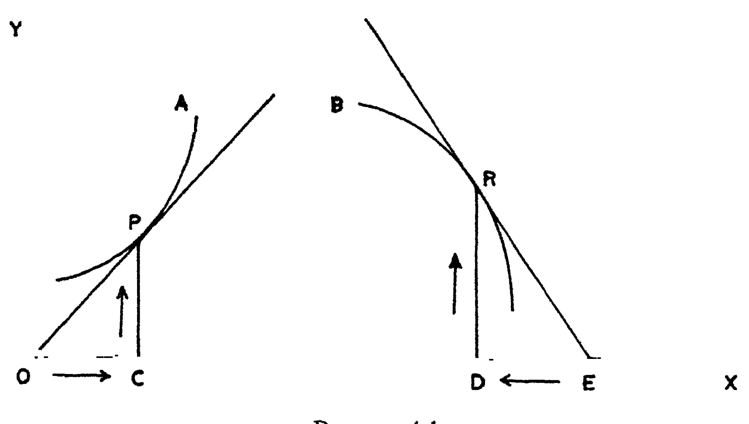


Diagram 4.1

The slope of the curve A at the point P is given by the tangent OP. This slope is positive because the trigonometrical tangent of the angle

POC made by the geometrical tangent with the axis of X is positive. And this trigonometrical tangent is positive because it is equal to CP/OP both of which are positive (they are both in positive directions).

The slope of the curve B at point R is given by the tangent RE. This slope is negative because the trigonometrical tangent of the angle RED made by the geometrical tangent with the axis of X is negative. And this is negative because it is equal to DR/ED of which ED is negative in direction.

Point of Inflexion

In Diagram 2.20 of Chapter II the point A on the total output curve is called the point of inflexion. It is so called because at this point the convexity (downward) of the curve changes into concavity. Whenever convexity changes into concavity or concavity into convexity we get a point of inflexion. We know that when a curve is convex it shows that y is increasing at an increasing rate as x increases. This means, in terms of differential calculus, that the derivative of the function is positive and is increasing. If it is increasing it means that the derivative of the derivative is positive. This was explained earlier. Similarly, when a curve is concave (as the part BC of the output curve) it shows that y is increasing at a decreasing rate as x increases. This, in terms of calculus, means that the derivative of the function is positive and is decreasing. If it is decreasing it means that the derivative of the derivative is negative, as was explained earlier.

At the point of inflexion, then, the derivative of the derivative, i.e., the second differential, changes from a positive to a negative magnitude. It is, therefore, zero at the point of inflexion. Taking the equation of the curve to be $y = Ax^3 + Bx^2 + Cx + D$, the first derivative of the function is $3Ax^2 + 2Bx + C$ and the second derivative is $6Ax + 2B$. If this is to be zero, x must be equal to $-B/3A$. And if this has to be positive (as it should be for a positive value of x) either B or A must be negative. As the output curve is first convex and then concave, the second differential should be positive first and then negative as explained above. Hence $6Ax + 2B$ should be first positive and then negative. It will be seen that the condition for this is that A should be negative and B positive.

CHAPTER 5

Electronic Computer

A COMPUTER is a machine that performs certain calculations at high speed. At one end, it is a calculator, such as the abacus, the simple adding and subtracting machine, and the electric calculator capable of the four simple arithmetic operations with or without a device for printing the results: at the other end, it is a digital computer (e.g., the Hollerith computer) and an electronic computer such as the ENIAC, the EDSAC, and the UNIVAC (all are commercial names of particular computers).

A computer is no substitute for the human mind. It cannot originate anything. It cannot do what we just order it to do; we must know how to order it if it has to do its work. If we ourselves do not know how to perform a certain calculation or analysis, we cannot tell a computer how to do it. A computer is just a tool which is effective in so far as the expert, with his skill, experience and ability, selects, directs and controls its use.

Operations of a Computer

A computer differs from a calculator as a gramophone does from a musical box. A gramophone can play any record placed on it while a musical box plays a single melody—the melody for which it is constructed. Similarly, a calculator has a built-in programme which cannot be changed at will. The steps that can be performed in a calculator are

limited by the programme of instructions already fed into it. The programme of instructions serves as a central automatic control. The required sequence of operations such as arithmetical operations, logical processes such as comparing and sorting, and holding within itself information fed while being worked upon—is preset. The machine issues the result after the operations are over. In a computer, the programme of instructions can be changed at will. Once a programme has been fed into it, it acts as a calculator. But given a different job, a new programme of instructions can be prepared and fed into the computer.

A computer is either a *digital computer* or an *electronic computer* (or *analog computer*). If we have paid some attention to the measurement of 'distance travelled' and 'speed' in a motor car, we can understand the difference between the two types easily. Distance travelled is measured digitally. Each digit on the distance-meter is provided with a separate indicator and we can increase the accuracy of measurement to any degree. In an electronic computer, machine numbers are represented by the strength of an electric current. While a digital computer works easily with ten to twelve digit numbers, an electronic computer operates with only a three or four digit precision.

The operations within a computer can be better understood by the information tabulated below with reference to industrial workers' wages:

	A Clerk	*Computer components*
1.	Receives reports regarding hours of work, absence, tax-allowances etc.	*Input*, which converts report into electrical impulses.
2.	Has a file of information on worker, e.g., worker's name, pay-scale, position etc.	*File*, which consists of data recordings on magnetic tape.
3.	Brings report and filed information together piece by piece in his mind, assisted at times by a scrap note-book	*Store*, which absorbs 'Input' and 'File' and holds information-data so long as required.
4.	Performs calculations to determine the pay.	*Arithmetic Unit*, which performs arithmetic and comparative operations.
5.	Uses a book of rules to regulate the entire process.	*Control*, which contains the PROGRAM of instructions carefully prepared and fed into the computer.
6.	Enters result on pay-sheets and pay-slips which are sent out.	*Output Unit*, which receives and prints results for distribution.

Errors and Limitations

A computer may produce incorrect results in three ways:

1. An error may be committed in recording or transcribing the results: so the input data may be wrong. Hence final results are bound to be wrong.
2. The *program* may be wrong either due to faulty analysis of

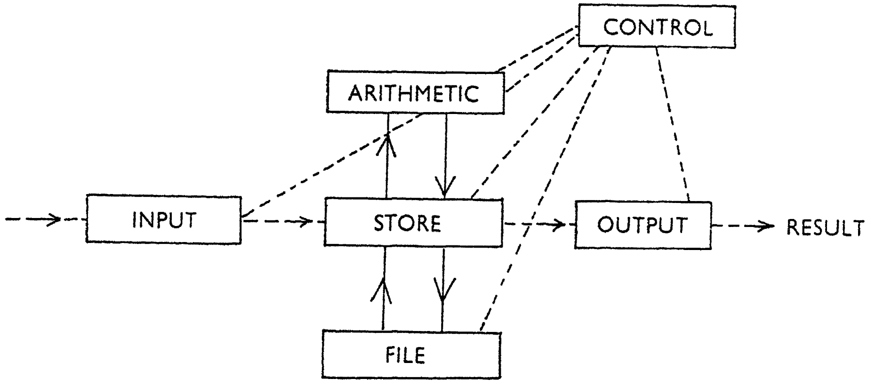


DIAGRAM 5.1 SHOWING COMPUTER COMPONENTS
(Dotted lines indicate overall control by the Program fed in)

the operations or due to errors of copying, transcribing and preparing it for feeding into the computer.

3. The computer may develop a fault (mechanical or electrical) and thus operations may become incorrect.

The first type of error is controlled by rechecking input data. The third source of error is controlled partly by usual accountancy checks and partly through built-in detecting-apparatuses. The analysis and testing of the second error is a very laborious and time-consuming process.

In common technical parlance the first type of error-system is called the G.I.G.O. (garbage in, garbage out) system. The control on the second type of error is called 'debugging' the *program*.

Four limitations of an electronic computer are mentioned below:

1. Partly because of the difficulty of preparing and feeding the *program* into the electric circuits, the accuracy of computation and results is low.
2. The computer lies idle while the program is being set, which takes a considerable time.
3. Each computing unit of the machine introduces errors. So the program has to be based on as few units as possible. On the one hand, this takes time: on the other hand, it diminishes the analogy with the real system.
4. The computing units are accurate only over a limited range.

COMPUTERS AND ECONOMICS

However, the electronic computer stands at the end of a long line of mechanical calculators. What distinguishes it is not any new magical

method of reasoning or calculation but the fact of its being automatic, fast and general purpose. Applied science has been calling for more and more calculations: techniques have been developed to meet this need. The computer is the result—a new and significant advance in the handling of data for science and economics (including industry and commerce). Its use is increasing in input-output analysis, linear programming, regressional analysis, operations research, game theory, econometrics, model-making and perspective planning.

Computers are coming into use in economics because, primarily, it is uneconomical to let experts "lose hours like slaves in the labour of calculations". The need for more and more calculations has been increasing owing to increasing centralisation of economic activities with wide coverage both in respect of region and persons involved (if not also, time !) and the increasing number of relationships written out in macroeconomic models.

The computer now enables an economist to take account of more complex and detailed economic models than ever before for the computer allows him to embark on calculations which he knows but which would take him years to do. For example, computer performs millions of calculations arising out of a large number of simultaneous equations at high speed and fairly accurately. But, since the computer limits the infinite mathematical operation at a point where the level of approximation is considered to be satisfactory, the final errors involved may become significant due to the mathematical steps being very numerous. Millions of operations cause another type of error—that due to the combined effect of electrical disturbances with which electronic computers are afflicted. These, together with implicit errors in the choice of an economic model and with the inaccuracy (due to measurement and sampling) of observations, limit the utility of electronic computers in economics.

PART TWO

Techniques of Analysis

CHAPTER 6

Macro-and Micro-Analyses

□□□□□□□□□□□□□□□□□

Analysis

THE SCIENTIFIC study of a subject consists in the establishment of causal relationships between phenomena. Our knowledge of facts has little use unless we are able to relate them one to another. The relationship with which a science is most concerned is the temporal relationship. We live our life spread over time and so the most important kind of knowledge is that which enables us to arrange phenomena on a time-scale. When they are thus arranged we are able to conceive of a phenomenon as the cause of that which follows it and effect of that which precedes it.

To establish such a relationship which, we repeat, is our concern in a science, we have to analyse facts, we have to break up a whole into parts before we can do anything else. To a casual observer facts and phenomena appear haphazardly mixed up so that the picture they present to the mind is absorbed by it as a consolidated whole. But such a *whole* means very little to the mind that absorbs *it*. It has to be broken up into its constituent parts which have later to be linked up to form a sequence of events. This process of breaking up and providing the needed links is called analysis.

MICRO- AND MACRO-ANALYSES

There are various kinds of analysis: we can break up a whole into big parts or into small parts. In economics, for example, we can break up the entire economy into large components, such as the consuming section and the producing section. We can then establish some relationship between them. Or we can break up the economy into small parts such as individual consumers and individual producers and then attempt to link them up with the help of causal links. When we break up an economy into big parts for purposes of analysis we are said to adopt the macrotechnique and the resulting economic theory is called macroeconomics. In the other case, our technique is said to be microtechnique and the theory resulting from such a study is called microeconomics.

The above statements do not, however, make the distinction between microeconomics and macroeconomics quite clear though they bring to light the fundamental difference between them. To understand the precise difference between the two techniques let us concentrate on the meanings of the words *micro* and *macro*. These words, borrowed from physical and biological sciences, mean *small* and *big* respectively. But a small thing is small only in comparison with a bigger thing. Hence, a thing that is small can be conceived of as forming a part of a bigger thing. Thus interpreted, micro can mean *a part* and macro *a whole*. And these interpretations are more suitable for purposes of economic analysis.

While in macroeconomics we are concerned with the study of the whole economy, in microeconomics we are concerned with the study of parts of an economy. The final object, in either case, is the same: we want to gain knowledge of the *whole* economy. In one case, we attempt to get knowledge by studying the whole, in its entirety; in the other, we try to acquire the needed knowledge by studying the whole piecemeal. There are advantages and disadvantages in studying an economy in its entirety. Economists before Keynes, particularly the neo-classical economists perfected, to a great extent, the microeconomic technique. All the advantages of such a technique had been secured. But then attention naturally came to be diverted to the other side and the shortcomings of such a technique began to attract our attention.

The greatest disadvantage of microtechnique revealed itself in connection with problems of economic fluctuations in general and with those of the relationship between factor-earnings and employment in particular. It was believed, for example, that employment of labour can be pushed up by lowering wages. Keynes, along with others, argued to show

that, if all the employers were to lower wages, less labourers would be employed than before. We shall explain this later on. Here we are concerned with noting the fact that microeconomic technique that had reached a state of great refinement during the neoclassical period has its shortcomings. And it is for this reason that later economists felt the necessity of applying a macroeconomic technique to the study of economic problems.

Micro- and Macro-Economies Defined

In the light of what has been said above, we can define a macroeconomy as one of which the environment is devoid of human influence, i.e., of the influence of other human, economic units. By contrast, a microeconomy is one which is subject to such influences emanating from its environment. Every economy, consisting of productive resources and consuming agencies, functions within an environment. In its efforts to satisfy its wants, it competes as well as cooperates with the forces that make up its environment. When we study a macroeconomy, the forces constituting the environment are natural forces only, as, by virtue of the economy being a macro one, all human elements are excluded from the framework within which it functions.

When, therefore, we study the economy of a country as a whole the only forces that it has to compete or cooperate with are those that are generated by the natural surroundings of the economy. This assumes, however, that the system under study is a closed one, i.e., there are no economic ties that bind it with other systems.

In the case of a microeconomy, the environment includes human influences also, so that the economic system has to compete or cooperate with other systems composed of human beings. The study of a firm, or an industry constitutes the study of a microeconomy for the reason that such a firm or an industry has to compete or cooperate with other firms or industries.

It is possible for us to make a macroeconomic study of an economy that is, in the physical sense, a microeconomy. This can be done by ignoring the influence of human elements in the environment. Macroeconomics can, therefore, be distinguished from macroeconomy: the former is a technique of study while the later is a physical system. In classical economics, a microeconomy was studied mostly with the use of macroeconomic technique. In doing this the individual unit studied was assumed to be free from human influences emanating from its environment. There are certain advantages to be gained by making a macroeconomic study of a microeconomy. The forces that operate from within a system—endogenous forces—can be studied with comparative ease when we ignore the impact on them of exogenous forces.

Difficulties of Microeconomic Analysis

When we study the economics of an individual unit we have to take into account all the forces that act on it. A firm, for example, is acted upon by the forces that emanate from other firms. The price that a firm can charge for its product does not merely depend on its cost of production and the demand of the consumers; it depends also on the prices that other firms charge for similar products. Besides this the price a firm can charge also depends on the income of the consumers. And the disposable income of consumers will depend, along with other things, on the prices of all other goods they buy. In short, the price of a particular commodity sold by a particular firm is a function of a large number of variables. And these variables are variables in the true sense of the word. They are all endogenous variables in the sense that a change in one of them causes a change in every other variable. A change in one variable causes and is caused by changes in all the other variables.

A microanalysis, therefore, has to take account of the complex interrelationships between the various micro units of a system. In this respect it is similar to the analysis of monopolistic competition among a small number of competitors. The decision taken by one of the monopolistic competitors depends on the decisions taken by all others. The action and reaction between one competitor and another introduce complications in the working of the system which can be avoided only by ignoring the interrelationships between them. In short, we can simplify the problem only by assuming that a particular unit under study is free from all inter-unital forces, particularly those that emanate from human agencies. When we make that assumption we, in fact, subject the unit to macroeconomic analysis.

The study of an economic unit made with the application of macroanalysis thus distorts the real picture of an economy. Nor can such a state of affairs be remedied by making a study of every unit individually and then aggregating the results obtained. For, in such a process of aggregation the errors too get aggregated. To apply the conclusions of a macroeconomic study of a microeconomy to the case of a macroeconomy is to commit a logical fallacy. Let us in this connection consider the relationship between wages and employment.

Wage-Employment Relationship

We have just said that to extend the conclusions of a microstudy to the case of a macroeconomy is to distort the true relationship between the

various economic entities. The classical and neoclassical economists believed that one of the ways of increasing employment was to reduce the wage rate. This belief was the outcome of microeconomic analysis of problems that are really macroeconomic in nature.

An individual employer of labour can employ a larger number of labourers if they are willing to work for a lower wage. And, therefore, when the money-wage is lowered employment increases, if other things remain the same. From such a conclusion, following from the consideration of the economy of an individual employer, one is likely to jump to the optimistic conclusion that, to remove unemployment of labour, wage rates must be reduced. It is forgotten, when applying the conclusion to the case of the whole economy, that it was assumed that other things remain the same. But other things do not in fact remain the same. When one employer offers a lower wage and succeeds in employing more labour the price of the product he sells is not materially altered. One important constituent of "other things" does remain the same for this employer. But when all the employers lower wage rates, the total income of the labourers might fall and, consequently, the demand for the goods produced goes down. It then would become unprofitable for the employers to maintain employment at the higher level.

Now it is not necessary that, when wage rates are lowered and more labourers are employed, the demand for the products must fall. But it is also not necessary that the demand must rise sufficiently to enable employers profitably to employ more labourers than before. The precise relationship between wage rates and employment would depend on the wage-elasticity of employment.

Keynes, speaking in terms of effective demand, maintained that when money-wage rates are lowered the effective demand for the products of labour might also get lowered so that the advantage of lower cost of production is neutralised by lower income. One's attention is drawn to such a relationship between the four variables of the system, namely, wage rates, cost of production, income and employment, when one adopts a macro point of view. In the case of an individual, wage rate, cost of production and employment appear to be the only variables to reckon with. Such a limitation of the number of variables is due to the fact that, when we apply microanalysis to the solution of a problem, we ignore the interrelationships of the various microunits of the system. The influence of a change in one variable on other variables escapes our attention. In the above example, the influence of a lower wage rate on the income of the employer remains unnoticed: the level of income is assumed to be a given datum or is treated as an exogenous variable. This is precisely what we observed earlier. We apply the macrotechnique to the solution of a microproblem. When we study the behaviour of an individual unit we cannot afford to ignore the interrelationships of the

various units. All the forces that are truly endogenous to the system must be taken account of.

ECONOMIC ENTITIES IN MACROECONOMICS

A macroeconomy is a closed economy in the sense that it is not open to external influences. The environments within which it functions do not contain any human element to act on the internal coherence of its endogenous forces. When we make a macroeconomic study of a community we have, therefore, to consider it as forming one unit and, as explained earlier, ignore the fact that its constituent elements act and react on one another. It is to get over the difficulties offered by such action and reaction that we prefer to treat a community of economic units as a single composite one and then study its behaviour. In so doing some useful information has perhaps to be surrendered; at any rate some knowledge of the manner in which the final adjustment of economic forces is arrived at has to be forgone. But in another direction we gain a good deal. We gain an overall picture of the economy in which the inter-unital forces are not ignored, though their precise nature is glossed over. For example, we do not take account of how the income and saving of one individual affect and are affected by the income and saving of other individuals, but we know how the national income and saving behave in a given set of circumstances.

In such a study, as everywhere else, we have to start with certain broad assumptions in regard to the aggregates of a system. Assumptions are the raw material of which economic theory is built. But these assumptions are not made at random; they are based on certain well-known properties of economic entities. Thus, for instance, we know that a community saves an increasing proportion of its income as income increases.

The economic entities we handle in the study of macroeconomics are those broad categories of variables that are relevant for the study and examination of the problems in hand. The focal point of interest for economists has, at least recently, been the size and variation of national income. We select, therefore, such variables as have a more or less direct relation to national income. We build a model out of variables that have a causal relationship to income. Hence, the variables of a system that we select for a macroeconomic study of a community are income, savings, investment and, therefore, expenditure. In all the models that are called macrodynamic models of an economy we make certain assumptions in regard to the behaviour of these aggregates of a system. In other words, we establish certain relationships between these variables. Expenditure, for instance, is taken to be a decreasing

function of income, and investment an increasing, constant or decreasing function of increment of income. In order to simplify the solution of our problem we at times make assumptions which are not strictly correct and then little by little, approach reality by progressively relinquishing these assumptions. We may, for example, start with the assumption that people save a constant proportion of their income and later give up this assumption.

In all macroeconomic studies our community behaves in precisely the same manner as Robinson Crusoe did. Not only are all consumers merged into one, even the consumers and producers are merged together. This merging is obvious to one who properly interprets the mathematical steps in the solution of a problem. Simultaneous solution of equations and the equations of savings and investment (ex-post and ex-ante) are, in effect devices for the merging of the production and consumption aspects, into a single whole. For all practical purposes, therefore, macroeconomics becomes Crusoe-economics.

Macroeconomics in Modern Studies

Classical economists were, in a way, interested in the study of macroeconomic problems. Their theory of distribution ran in macro terms. They explained how the total wealth produced was distributed among the three classes of income-producers. While it is clear to us today that their theories were defective it cannot be denied that their concern was with problems that we recognise as macroeconomic today. Further, they introduced a dynamic element into the same study by discussing such questions as that of profit tending to become zero in the long run. Even here they were vague; their concept of profit was analytically wrong and their notions of the *long-run* was historical rather than analytical and ill-fitted into the logical framework of economic theory.

The neoclassical economists were concerned with the task of refinement of theory. They focussed their attention, therefore, on the decisions of individual economic units. The economy of a country functions through the decisions of individual producers and individual consumers. Neoclassical economists perfected an analysis to deal with the problems of individual buyers and sellers, of individual consumers and producers. Such a step was necessary; the neoclassical theoretical structure helped our understanding of the interplay of economic forces and prepared the ground for the extension of the same analysis to cases of entire economies.

But it is difficult to expect that the human mind will observe strict discipline in matters of thinking. While the neoclassical writers were thus concerned with microeconomic problems they allowed macroeconomic considerations to make an encroachment upon their study.

There were pressing applied economic problems to which they could not shut their eyes. Income and employment were fluctuating and they were interested in knowing why they were not constant. Something was needed to formulate a suitable economic policy to secure the stability of a system. Bohm-Bawerk, Schumpeter, Marshall and Pigou allowed themselves to take up macroeconomic problems. National dividend, social welfare, economic evolution and the trade cycle are concepts that are essentially macro in their nature and it is with them that the neoclassical economists were concerned in the main body of their economic theory. But old habits die hard and it took time for economics to put itself soundly on a macroeconomic foundation. Keynes took a bold step in that direction and his economics as expounded in his *General Theory* is, in the main, a study in terms of macro concepts of economic science. Since his time, it has become almost a fashion to talk in terms of what are called the aggregates of an economic system. National income rather than the income of an individual producer, national expenditure rather than the expenditure of an individual household or a firm and, similarly, national saving and national investment constituted the entities in terms of which models of an economy began to be constructed. The step was, from one point of view, in the right direction. Our preoccupation with the conditions necessary for the equilibrium of individual units of a system made us forget that those conditions operated within the general framework of a broader set of conditions for the equilibrium of the whole system.

But no innovator's work can be complete; it has to depend on the efforts of those who follow to give it a finish. Keynes's theory has been criticised and the post-Keynesians have done some useful work in rounding off the various angularities in the *General Theory*. But perhaps some of the post-Keynesians are as Keynesian as Keynes himself was. That is but natural; those who are post-Keynesian in the historical sense are not always post-Keynesian in the analytical sense. But that is a point with which we are not concerned here.

Keynes's Macroeconomics Criticised

Is Keynes's theory completely macroeconomic? Is it macrostatic or macrodynamic? Is it a *General* theory and, if so, in what precise sense? These and similar questions are meant to throw doubts on the soundness of Keynes's economic theory. But it is not questions of this nature that we are concerned with here. Keynes's macroeconomics runs, says Professor Boulding, in terms of flows rather than in terms of stocks. The main body of his theory is concerned with the circulation or exchange of assets rather than their creation and destruction. Such a deficiency

in Keynes's analysis does not make it less macroeconomic but it does restrict its usefulness. Boulding has, as we know, introduced the balance-sheet approach in economics. An economic unit is not merely concerned with what happens to its income (flow), it is also interested in what happens to its capital (stock). This inclusion of capital-stock in the calculus of an operating unit reminds us of the controversy of *flow* versus *stock* approach in economic theory. Whether Boulding's criticism of Keynes's work is valid or not might be beside the point here but it serves to limelight the importance of capital-considerations in economics.

Again and again the importance of the capital aspect of economic entities asserts itself. Is money a stock (capital) concept or a flow concept? Is a firm interested in maximising its net income or (the value of) its capital stock? Should we, accordingly, include capital stocks in our economic models or let them run in terms of incomes or flows? Should we take account of what happens to the production capacity of an economy when investment is increased or confine our attention only to the income-creating effect of investment? Boulding's insistence on the inclusion of capital considerations in economic theory is in line with these developments.

Keynes's macroeconomics fails to provide a theory of distribution. Boulding says in this context that "Another weakness of the Keynesian economics is its failure to provide any 'macroeconomic' theory of distribution commensurate with its theory of employment", and then offers us a macroeconomic theory of distribution which in its general pattern resembles the classical theory.

MACROECONOMICS ESSENTIALLY CRUSOE-ECONOMICS

In macroeconomics our universe of study is free from the exogenous influences emanating from human beings. A macroeconomy operates in an environment that consists of natural forces only. If such an environment is hard to find we just ignore the human forces in the environments. It is for this reason that macroeconomics becomes, for all purposes, Robinson Crusoe-economics. All the advantages of macroeconomic study can, therefore, be claimed for the study of Crusoe-economics. What we attempt to do in all macrostudies is to consider all the units of a factor together: we treat them as forming one consolidated, immutable entity. And the best way to do that successfully is to study the economics of a single individual. In the case of Crusoe, not only are all the units of a factor consolidated, even the various factors of production are, as it were, lumped together. One individual represents all the units of all the factors: macroeconomics finds its culmination in the economics of Robinson Crusoe.

In macroeconomics we are concerned with the determination of, what are called, the values of aggregates. We have for instance to determine the total or aggregate savings, total investment, total income, total capital stocks and so on. Likewise, we should determine, as Boulding says, the total share of labourers, the total share of capitalists and the total share of entrepreneurs. When we study Crusoe economy we find a more thorough blending of all these entities. It is difficult to distinguish savings from investment (though it is not impossible to do so), income from total spending, wages from other distributive shares and so on. But in the measure in which, in the study of Crusoe's economy, it is difficult to make such distinctions, in that measure is it easy to perfect the macroeconomic technique.

Macroeconomics differs from microeconomics in as much as the environments within which the decision-taking units operate in the two cases are different. In the case of microeconomics, there is inter-unital competition which, in macroeconomics, it is possible to assume away. In the case of a Crusoe-economy the macroeconomic technique finds itself so perfected that inter-unital competition actually ceases to exist.

All our propositions in macroeconomics should, therefore, be applicable with vengeance to the case of Robinson Crusoe. Nay, even our microeconomic theories can be tested in the test-tube of Robinson Crusoe. For, economics in the comprehensiveness of its discipline must yield propositions that have application to individuals as much as to groups of individuals.

CHAPTER 7

Static and Dynamic Analyses

MEANING

IN ECONOMICS, static analysis refers to that part of a study in which phenomena and situations as well as effects of changes therein are analysed without reference to time. Thus when we say that

(1) if price is lowered by 10 percent, demand rises by 7 per cent, or
(2) if a commodity-tax is levied, the price will increase but not by the full amount of the tax,

we are in the field of static analysis.

Dynamic analysis refers to that part of the study of economics in which either time appears as a factor explicitly or in which various factors (whether causal or caused) are mentioned with reference to time. Thus when we say that

(1) last year's price of wheat determines the current year's acreage under wheat, or,
(2) population is growing at two per cent per year, or,
(3) demand for wheat is determined by price, income and time, or
(4) production depends on labour and capital and time, or,
(5) current investment is determined by the change in income experienced by the people last year,

we are in the field of dynamic analysis. Symbolically, the last two statements may be written as follows:

(*a*) $x = a.L^b.C^c.d^t$, where a, b, c and d are constants and x, L, C and t stand respectively for output, labour, capital and time;

(b) $I_t = a + b(Y_t - Y_{t-1})$, where a and b are constants and I_t is the current year's investment expressed in terms of the current year's income, Y_t, and last year's income, Y_{t-1}.

In the first relationship time appears explicitly: in the second relationship, investment and incomes are referred to with time as a suffix.

Dynamics, Time and Time-suffixes

The terms 'static' and 'dynamic' are borrowed from mathematics. There also, 'statics' refers to a study of final (equilibrium) situations without reference to time; and 'dynamics', to a study where time comes in as an explicit factor. In simple cases, it is easy to show that a relationship with such a use of time can be restated as a relationship between variables with time-suffixes. Thus,

$$Y = a + bt + ct^2,$$

is the same thing as

$$Y_t = a + bt + ct^2.$$

We can also write

$$Y_{t-1} = a + b(t-1) + c(t-1)^2$$
$$Y_{t-2} = a + b(t-2) + c(t-2)^2$$

Hence

$$Y_t - Y_{t-1} = b + 2ct - c$$

and $$Y_{t-1} - Y_{t-2} = b + 2c(t-1) - c$$

so that $$(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = 2c$$

which is a relationship among variables with time-suffixes.

Spatial Dynamics

We can extend the meaning of dynamic analysis in another sense. We may express the relationship between phenomena observed at different places. Thus, we may say that

(1) the marketed quantity of wheat in urban areas increases when wheat produced in rural areas increases in quantity;
(2) rise in the wage demanded by privately employed labour depends on the rise allowed to those employed by the government; and
(3) a rise in consumption by X leads to a rise in the consumption by his neighbour Y as also by his office-employee, Z.

These may be said to involve *spatial dynamics*. In a sense, a change in space is equivalent to a change of time and just as we showed above that

$$Y = a + bt + ct^2$$

can be expressed as

$$(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = 2c,$$

a relationship involving space-factors can be converted into one involving factors with time-suffixes. But, in practice, certain observations and data are more convenient and easy to get. Hence, instead of converting them into forms with only time-suffixes, they are used as such. That is why, starting from a few fundamental factors of production in economics, we now consider dozens of inputs in the name of input-output analysis.

Incidentally, of late, in economics, dynamic analysis has been associated with the introduction of expectations in an analysis. An 'expectation' is at base a foretold measure of a phenomenon; and any foretelling is mainly a function of past observations. Instead of bringing these past factors into the analysis with time-suffixes, it is more convenient to introduce the factor 'expectation' without its determinant function. An expectation is sometimes expressed explicitly in terms of factors with time-suffixes.

Why Dynamic Analysis?

Since reality is inevitably associated with time, and since economics is a study to help understand reality and even to influence future reality, dynamic analysis is relevant. In recent years, such analysis has been undertaken to elucidate the process of development and how (and to what extent) its determinants can be controlled.*

As we shall see later, it is helpful to make a long-term dynamic analysis first for a better short-term dynamic analysis. It is however not correct to say that it alone is relevant, useful and desirable. For at the back of the intention to understand and influence reality stands the desire to escape pain, dissatisfaction and fluctuations. And, although, in a sense, matter is limitless, relative to the limited life of an individual, matter and resources—factors and goods—are limited in stock and, hence, their flows are also limited. But there is a tendency to make use of greater and greater quantities of these items. This leads to shortage of supplies and over-use of factors resulting in inefficiencies, shortages and fluctuations. Unless, therefore, we limit our demands (and hence, wants), fluctuation, dissatisfaction and pain are inevitable. The desire to escape pain, dissatisfaction and fluctuations can be reinterpreted as the desire to attain a state of constancy—a static equilibrium. Hence it is natural that a study (or analysis) should be made in terms of factors without time-suffixes.

*Politically, we think, dynamic analysis has been undertaken to establish the desirability of a laissez faire or liberal economic policy.

Static Analysis

Static analysis provides an end-view of factors in operation and, in a sense, the end is more important than what happens along the path to the end. There is yet much unexplored ground to be covered—which is not only relevant and useful in the current context of worldly life but which is desirable for keeping us thinking and imagining. We have yet to study production functions for many goods—both industrial and particularly agricultural—and services in terms of inputs, as also variables measuring the quality of inputs. Similarly, objective social welfare functions, with or without a measure of the degree of decentralisation, are needed in a form more appropriate for planning.

Again, there is scope for development of static input-output analysis with regard to internal and international trade as also distribution of national income consequent on the availability of (*i*) more detailed and comprehensive data; as also (*ii*) know-how from mathematics and statistics. The engineering approach to the study of economic systems deserves to be worked in practice.

Likewise, static analysis is being increasingly made (and it deserves promotion) of the institutional determinants of an economic system.

In view of the increasing extent, particularly in the background of backward but developing economies, to which individual preferences are acquired, taught (through advertisement) and imposed, the assumption that preferences are innate may be reconsidered. Similarly, there seems to be a growing conflict between individual preferences and social preferences.

Analysis of Static Production Function

In economics it is assumed that factors of production are paid according to marginal productivity and that this would (or should?) exhaust the output, and that the factors yield decreasing returns ultimately. Let us assume that there are two factors of production (labour, L, and capital, C) and work with the production function:

$$x = a + bL + cC$$

Then, on increasing labour by one unit we get (say) x_{1l} and

$$x_{1l} = a + b\,(L + 1) + cC$$

and hence the marginal productivity of labour is

$$x_{1l} - x = b$$

Similarly, if by increasing one unit of capital we get output x_{1c}, we shall have

$$x_{1c} - x = c$$

Hence the total payment at output x-stage will be b times L and c times C. Therefore the surplus left after payment will be

$$x - (bL + cC) \text{ or, merely } a$$

Now, a is inadmissible, because that would mean that even if no labour and capital are employed (i.e., if $L = 0$ and $C = 0$), the output will be a. So a must be zero. Hence, there will be no surplus left: the output will be exhausted. However, this does not ensure any decreasing returns for any factor: the marginal productivity is constant. Hence we change to a new production function

$$x = b_1L - b_2L^2 + cC$$

and then

$$x_{1l} - x = b_1 - 2b_2L - b_2$$
$$= (b_1 - b_2) - 2b_2L$$

which means that as the quantity of labour (L) increases the marginal productivity decreases. In this case, the marginal productivity of capital (C) remains c as before. The total payment to factors of production according to marginal productivity will be

$$\{(b_1 - b_2) - 2b_2L\} \text{ times } L \text{ plus } c \text{ times } C$$

and the surplus left after payment will be

$$(b_1L - b_2L^2 + cC) - \{(b_1 - b_2)\,L - 2b_2L^2 + cC\}$$

or,

$$b_2L + b_2L^2$$

This means that the greater the amount of labour used and the greater the degree of diminishing returns, the greater the surplus that will be left after payment to factors of production. If there is no social arrangement for proper mobilisation and distribution of this surplus it will lead to an uneven distribution of income: for it is very likely to be pocketed by those in power, whether political or economic.

Again, if we introduce into the production function an element of increasing returns for capital by assuming

$$x = b_1L - b_2L^2 + c_1C + c_2C^2$$

then the marginal productivities shall be given by

$$x_{1l} - x = (b_1 - b_2) - 2b_2L$$

and

$$x_{1c} - x = (c_1 + c_2) + 2c_2C$$

Hence the surplus left after payment will be given by

$$(b_2L + b_2L^2) - (c_2C + c_2C^2)$$

which means that the surplus may be less or more according to the degree of returns with respect to labour and capital as also their quantities. Due to the existence of a negative sign between the surplus due to labour and that due to capital, the surplus may even be zero.

Since there is no mention of time, this is a static analysis. It brings out the fact that with such production functions and assuming payment to factors according to their marginal productivity, the output is not likely to get exhausted. Even if it were exhausted, the question would remain whether such distribution of national product is in keeping with any desired social welfare function.

Cobb-Douglas Production Function

To this state of affairs, there may be added the empirical experience from western studies that the share of labour in national income is more or less constant. There has therefore been a shift to another production function, a Cobb-Douglas production function of degree one:

$$x = aL^{b}C^{1-b}$$

where the marginal productivity of labour is bx/L and that of capital, $(1 - b)\,x/C$, so that an increase in the quantity of labour at constant output leads to a decrease in its marginal productivity. This applies to capital also: it is not possible to have such combination as "decreasing returns for labour and increasing returns for capital". However, the total payment will now be

$$\begin{aligned} &(bx/L).L + (1 - b)\,(x/C).C \\ &= bx + (1 - b\,)x \\ &= x \end{aligned}$$

Hence the surplus left will be zero: payment to the factors will exhaust the output.

The Cobb-Douglas production function is also a constant elasticity of substitution function (CES function). There are other CES functions which have been used. An attempt has also been made to introduce a time element in this function to make it dynamic, but then it modifies the empirical proposition of product exhaustion and constant share of labour in national product.

However, such static analysis does help to bring out long period implications and indicate policy-directions. Thus, in a model where (*i*) wages (w) are a function of the demand for labour, (*ii*) employment (e) or demand for labour depends on national output (Y) and (*iii*) the propensities to spend for labour and others are respectively k_1 and k_2, we may have

$$w = f_1(e) \qquad \ldots(1)$$

$$e = f_2(Y) \qquad \ldots(2)$$

$$\text{Labour's expense} = k_1.w.e$$

$$\text{Others' expense} = k_2.(Y - w.e)$$

And since these two expenses would exhaust the output,

$$Y = k_1.w.e + k_2 (Y - w.e) \qquad \ldots(3)$$

Using the three equations, one can determine w, e and Y. The value of Y will be the equilibrium value and $w.e/Y$ will give the share of labour in national income. Since employment (e) is determined independently of wages (w), if w is slightly increased by w', the total expense (and, hence, total demand) will go up by

$$k_1.w'e - k_2.w'e$$

which will be positive if k_1 is greater than k_2, i.e., if the labourers' propensity to spend is greater than that of the 'others'. This should increase (*i*) the share of labour as also (*ii*) employment. So it suggests that during a depression it may be useful to pay more wages and thus help send up Y and revive the economy. This is just an indication which may be helpful: strictly speaking, higher wages may lead to the invention and adoption of labour-saving devices. However, such static analysis does help.

Static Analysis and Optimal Allocation

Similarly, in the discussion of best or optimal equilibrium, static analysis holds out suggestive lines (*i*) without making it explicit how necessary information for following the suggested lines can be marshalled, and (*ii*) without giving any help in regard to the distribution problem. It is easy to recall that static analysis suggests that—

1. Two commodities should be distributed between two consumers such that the marginal rates of substitution of one commodity for another is the same for each consumer. Otherwise, the two consumers will gain by mutual exchange.
2. Two inputs should be allocated between two production lines such that the ratio of the marginal productivities of the two inputs is the same for each line of production. Otherwise, more can result by reallocating the inputs.
3. Two commodities should be produced to the extent that the ratio of the marginal social utility to marginal social cost is the same for each commodity. Otherwise, it will benefit society to produce more of that commodity for which the ratio is higher.

Of course static analysis will not help us (*i*) to decide what should be meant by 'social utility' and 'social cost' and (*ii*) to measure them. According to the second suggestion, use should be made of the expression 'marginal social benefit' or 'marginal social utility' instead of 'marginal productivities': but static analysis cannot help us to know how to measure the same. Or, again, according to the first suggestion, we

have to allow the total utility function of a consumer to be maximised while leaving the total utility of the other consumer unchanged.

These rules flow from static analysis, using the arguments one is accustomed to in the study of the law of substitution. For, at given prices, with free exchange and free mobility, each consumer maximises his total utility from (say) two commodities by making

$$\frac{\text{Marginal utility of one commodity}}{\text{Marginal utility of other commodity}} = \frac{\text{Price of first commodity}}{\text{Price of second commodity}}$$

and each producer maximises his profit by making

$$\frac{\text{Marginal productivity of one factor}}{\text{Marginal productivity of other factor}} = \frac{\text{Price of first factor}}{\text{Price of second factor}}$$

And from

$$\begin{aligned}\text{Marginal utility} &= \text{price (according to consumer)}\\ &= \text{marginal cost (according to producer)},\end{aligned}$$

we get a constancy of the ratio of marginal benefit to marginal cost.

But then the marginal utility to an individual need not be the same thing as marginal utility to society: and marginal cost to a firm need not be the same as the marginal social cost. This implies producing more goods for those who can pay for them than for those who cannot. If we ignore these remarks, we can use this static analysis to conclude that an equilibrium attained under perfect competition is conducive to maximum welfare. In contrast, resources will be used less in monopoly and monopolistic industries. But what about an economic system where there is only monopolistic production? And what about static analysis for an economic system with centralised planning? Well, we have a whole lot of static analysis not only in the directions we have touched upon but also in those we have not. Such analysis, it may be repeated, gives us a sense of direction and values and enables us to build up and improve dynamic analysis.

Those who are interested only in static analysis are (perhaps) rather wantless, more undisturbed about how things shape themselves through time and more prepared to face the interim events with a smile, keeping their eyes focussed on the final static, stable goal. But the world and the common man are not made that way. We experience fluctuations, now irregular, now cyclical, and we see that some countries seem to be going up and up the ladder of economic prosperity notwithstanding their philosophy of multiplicity of (and actually increasing) wants as also up the ladder of economic and political power, accompanied by growing knowledge, unravelling of Nature's secrets, conquest over Nature and flights into space. Besides, there are the developing countries with their problems of marketed agricultural surplus, foreign trade and balance of payments. These call for a dynamic analysis. Incidentally, earlier there had been curiosity about the dynamics of sensitive market demand

and sensitive market supply involving the cobweb theorem—the regular, damping or exploding cycles over time.

Dynamic Analysis

Dynamic analysis may be classified as dealing with (*i*) trends and (*ii*) short-term movements (such as trade cycles, crop-fluctuations leading to speculation, etc.). Short-term dynamic analysis is generally based on assumptions of constant technology, constant capital and constant population. Long-term (trend-) dynamic analysis is meant to trace mainly the growth of national output and the influence of factors assumed constant for short-term analysis: but it is also used to study the repercussions of foreign trade, propensity to import, the theory of savings and the theory of intended investments. Thus, as we shall see in studies of economic models and in the study of the multiplier and acceleration in this book, the growth of national income is seen as dependent on various combinations of such alternatives as

$$\text{Savings} = \text{(i) } s.Y_t, \text{ or,}$$
$$= \text{(ii) } Y_t - C_t = Y_t - c.Y_{t-1}$$

and

$$\text{Intended Investment} = \text{(i) } g(Y_t - Y_{t-1}), \text{ or}$$
$$= \text{(ii) } g(C_t - C_{t-1}), \text{ or } gc(Y_{t-1} - Y_{t-2}), \text{ or,}$$
$$= \text{(iii) } g(Y_{t-1} - Y_{t-2}) + G$$

where Y stands for income; C for consumption and G for Government investment, while c, s and g are constant propensities.

At the back of attempts to make such dynamic analysis, the idea is to know how an economy behaved in the past, how it may be expected to behave in the future and how (and to what extent) it can be controlled.

Long-term Dynamic Analysis

To understand the growth of output long-term dynamic analysis has been attempted by drawing freehand curves through points plotted on graph paper or by calculating moving averages, simple and weighted, but these do not help us much in causal analysis and are difficult to connect with any economic theory of development. So, particular curves have been selected and fitted:

1. *Straight line*—particularly for short intervals and as a first approximation to different curves. Where compound or geometrical rate of growth is considered apt in economic theory, a straight line fitted to the

logarithms is also a good approximation:

$$x = a.L^b.C^{(1-b)}$$

or, $$\log x = \log a + b \log L + (1 - b) \log C$$

which is a straight line relationship between $\log x$, $\log L$ and $\log C$.

2. *Parabola*, e.g., $x = a + bt + ct^2$
3. *Exponential curve*, e.g., $x = a.e^{bt}$
4. *Logistic curve*, e.g., $x = \dfrac{k}{1 + ae^{-bt}}$

which is justified wherever the rate of growth is held by theory to be proportional to (*i*) output (x); and (*ii*) the gap, $k—x$, yet to be bridged, k being the maximum attainable output.

5. *Gompertz curve*, $x = a.e^{-bc^t}$,

which also implies a maximum attainable output.

6. *Cobb-Douglas curve* with an efficiency element,

$$x = (1 + a)^t.L^b.C^{(1-b)}$$

where L and C are labour and capital and t stands for time: over time the efficiency increases at the rate a. This does imply constant returns to scale, which may be justified within narrow limits in the context of national output. The function has been verified more for individual industries (and is used even in connection with agricultural production) than for national output. The function also implies that labour and capital are complete substitutes, though we know they may be—in general, are—complementary to an extent.

Long-term trend studies presuppose that the short-term movements round it are neither purely cyclical nor explosive but are damped in nature. For with explosive short-term fluctuations the utility of trend studies is greatly reduced.

It has been attempted to build a complete economic model round some of these trend equations. Alternatively, economic models have been built round savings and investment equations, or round the multiplier and the accelerator: a good many of these have been mentioned in connection with acceleration. Nevertheless we give one below:

A Dynamic Model

We assume that exports (X) are increasing exponentially; imports (M) have a floor and are proportional to income; consumption (C) has a floor and is also proportional to income; and investment (I) is proportional to rate of change of income $\left(\dfrac{dY}{dt}\right)$. The balance of payments (B) is given by

$$B = p_x.X - p_m.M$$

$$\text{Income, } Y = C + I + X - M$$

The other conditions may be quantified as follows:

$$X = X_o.e^{at},$$

where X_o is "initial exports" and t stands for time: exports increase geometrically.

$$M = m + m_i.Y$$

$$C = c_o + c.Y$$

$$I = b.\frac{dY}{dt}$$

Substituting the values of C, I, X and M in

$$Y = C + I + X - M$$

and simplifying we get

$$\frac{dY}{dt} - \left(\frac{1 - c + m_i}{b}\right). Y = (-X_o/b).e^{at} + \frac{m - c_o}{b}$$

Putting $(m - c_o)/b = \lambda_o$; $-X_o/b = \lambda_1$; and $(1 - c + m_i)/b = \lambda_2$,

we get
$$\frac{dY}{dt} - \lambda_2 Y = \lambda_1 e^{at} + \lambda_o$$

The solution of this differential equation is given by

$$Ye^{-\lambda_2 t} = \frac{\lambda_1}{a - \lambda_2} e^{(a-\lambda_2)t} - \frac{\lambda_0}{\lambda_2} e^{-\lambda_2 t} + \text{constant}$$

Note: (i) If $\lambda_2 = a$, the solution will reduce to

$$Ye^{-\lambda_2 t} = \lambda_1 t - \frac{\lambda_0}{\lambda_2} e^{-\lambda_2 t} + \text{constant}$$

(ii) If, further $\lambda_0 = 0$, i.e., if $m = c_0$, then

$$Ye^{-\lambda_2 t} = \lambda_{1t} + \text{constant}$$

Now, if at $t = 0$, $Y = Y_0$, then the constant is given by

$$\text{Constant} = Y_0 - \frac{\lambda_1}{a - \lambda_2} + \frac{\lambda_0}{\lambda_2}$$

$$\therefore \quad Y = \left\{Y_0 - \frac{\lambda_1}{a - \lambda_2} + \frac{\lambda_0}{\lambda_2}\right\} e^{\lambda_2 t} + \frac{\lambda_1}{a - \lambda_2} e^{at} - \frac{\lambda_0}{\lambda_2}$$

$$\because \quad \frac{dY}{dt} - \lambda_2 Y = \lambda_1 e^{at} + \lambda_0$$

$$\therefore \quad \frac{1}{Y}\frac{dY}{dt} = Y = \lambda_2 + \frac{\lambda_1 e^{at} + \lambda_0}{Y}$$

which is the expression for the rate (Y) of growth of Y. If we replace Y (on the right hand side) by the expression for Y in terms of t, we get

the growth rate as a function of t.

This is not a simple function. So we see that even a simple attempt to approach reality leads us to a solution which is mathematically complex and difficult to interpret economically.

Nevertheless, we attempt below an analysis—dynamic of course—to find a place for technical progress in growth.

Technical Progress in Dynamic Analysis

It is difficult to define technical progress: it is due to technical progress that, with a fully employed given combination of inputs, the yield increases. Technical progress does not raise the yield of all inputs to the same extent: nor does it show itself up in every field and area. How to include it in an income equation is a problem. In some cases, in a static Cobb-Douglas Production Function, a multiplying factor, a^t, is added:

$$Y = a^t.L^b.C^{(1-b)}$$

We propose to work with the following model:

$$Y = f_1(t).f_2(L,C) = f_1 f_2 \text{ (in brief)}$$

where $f_1(t)$ is a function of time implying a steady rise in national income owing to technical progress. It is independent of inputs—here, labour, L, and capital, C: this is less realistic as technical progress is seldom unaffected by the existing state of labour and capital resources. However, taking logarithms, we get,

$$\log Y = \log f_1 + \log f_2$$

Differentiating with respect to time, t, we get

$$\dot{Y} = \frac{1}{f_1}\cdot f_1' + \frac{1}{f_2}\left[\frac{\delta f_2}{\delta L}\cdot\frac{dL}{dt} + \frac{\delta f_2}{\delta C}\cdot\frac{dC}{dt}\right]$$

$$= \dot{f_1} + \frac{1}{f_1 \cdot f_2}\left\{L.\frac{\delta(f_1 \cdot f_2)}{\delta L}\cdot\left(\frac{1}{L}\frac{dL}{dt}\right) + C.\frac{\delta(f_1 \cdot f_2)}{\delta C}\cdot\left(\frac{1}{C}\cdot\frac{dC}{dt}\right)\right\}$$

$$= \dot{f_1} + \left(\frac{L}{Y}\frac{\delta Y}{\delta L}\right)\left(\frac{1}{L}\frac{dL}{dt}\right) + \left(\frac{C}{Y}\cdot\frac{\delta Y}{\delta C}\right)\left(\frac{1}{C}\cdot\frac{dC}{dt}\right)$$

Since elasticity of production with reference to labour will be

$$\delta Y/Y \div \delta L/L = \text{(say) } \xi_l$$

and that with reference to capital,

$$\delta Y/Y \div \delta C/C = \text{(say) } \xi_c$$

and since, like $\dot{Y}$, the growth rate of technical progress, labour and capital are given by $\dot{f_1}$, $\dot{L}$ and $\dot{C}$, we can write the above expression as follows:

$$\dot{Y} = \dot{f_1} + \xi_l.\dot{L} + \xi_c.\dot{C}$$

Let Y_l and Y_c indicate the marginal productivity of labour and capital. Then, assuming constant returns to scale, we have, by Euler's theorem, that when the inputs are paid according to marginal productivities the product is exhausted, i.e.,

$$Y = Y_l.L + Y_c.C$$

Differentiating with respect to time, t, we get

$$\frac{dY}{dt} = Y_l.\frac{dL}{dt} + L.\frac{dY_l}{dt} + Y_c.\frac{dC}{dt} + C.\frac{dY_c}{dt}$$

Dividing both sides by Y and simplifying in terms of $\dot{Y}$, $\dot{Y}_l$, Y_c, $\dot{L}$, $\dot{C}$, ξ_l, and, ξ_c, we get

$$\dot{Y} = \xi_l.\dot{L} + \xi_l.\dot{Y}_l + \xi_c.\dot{Y}_c + \xi_c.\dot{C}$$

but it is already proved that

$$\dot{Y} = \dot{f}_1 + \xi_l.\dot{L} + \xi_c.\dot{C}$$

Equating the two expressions we get

$$\xi_l.\dot{Y}_l + \xi_c.\dot{Y}_c = \dot{f}_1$$

It may now be argued that even if diminishing returns operate to scale, i.e., even if ξ_l and ξ_c be together less than unity, $\dot{f}_1$, rate of technical progress, may be large enough to justify (*i*) positive values of $\dot{Y}_l$ and $\dot{Y}_c$ so that both Y_l and Y_c may increase, as also (*ii*) a positive value of $\dot{Y}_l$ even if $\dot{Y}_c$ is zero or negative so that Y_l may be increasing even if Y_c is stationary or falling. So, the higher the technical progress, the greater may be the rate of growth of income to labour and capital.

Again, since we know that, by definition,

$$\xi_l = \frac{L}{Y}\cdot Y_l$$

$$Y_l = \frac{Y}{L}\cdot \xi_l$$

and taking logarithms,

$$\log Y_l = \log Y + \log \xi_l - \log L$$

Hence, on differentiation with reference to time, t, we get

$$\begin{aligned}\dot{Y}_l &= \dot{Y} + \dot{\xi}_l - L \\ &= (\dot{f}_1 + \xi_l.\dot{L} + \xi_c.C) + \dot{\xi}_l - \dot{L} \\ &= \dot{\xi}_l + \dot{f}_1 - (1 - \xi_l)\,\dot{L} + \xi_c.\dot{C}\end{aligned}$$

which, if there be constant returns to scale, i.e., if $\xi_l + \xi_c = 1$,reduces to

$$\begin{aligned}\dot{Y}_l &= \dot{\xi}_l + \dot{f}_1 - \xi_c.\dot{L} + \xi_c.\dot{C} \\ &= \dot{\xi}_l + \dot{f}_1 + \xi_c\,(\dot{C} - \dot{L})\end{aligned}$$

from which it can be concluded that, if we ignore the first two terms on the right hand side as being too small, the marginal productivity of

labour will increase if capital grows faster than labour. Since wages are equal to marginal productivity, it can be said that the real wages of labour will increase with time if capital increases faster than labour.

If we want that per capita income should not decrease even if diminishing returns to scale operate, we may proceed from

$$\dot{Y} = f_1 + \xi_l.\dot{L} + \xi_c.\dot{C}$$

and put $\xi_l + \xi_c = 1 - \varepsilon$, where ε is a positive number less than one. Then if the per capita income is not to decrease, $\dot{Y} = \dot{L}$ and

$$\therefore \quad \dot{L} = \dot{f_1} + \xi_l.\dot{L} + \xi_c.\dot{C}$$

or,

$$(1 - \xi_l)\,\dot{L} = f_1 + \xi_c.\dot{C}$$

or,

$$\dot{C} = \left(\frac{1 - \xi_l}{\xi_c}\right).\dot{L} - \frac{\dot{f_1}}{\xi_c}$$

$$= \frac{\xi_c + \varepsilon}{\xi_c} \cdot \dot{L} - \frac{\dot{f_1}}{\xi_c}$$

$$= \left(1 + \frac{\varepsilon}{\xi_c}\right) \cdot \dot{L} - \frac{\dot{f_1}}{\xi_c} \text{ or, } \dot{L} - \frac{1}{\xi_c}(\dot{f_1} - \varepsilon\,\dot{L})$$

So, if there be no technical progress, capital must increase faster than labour in order that the per capita income may remain the same. In the presence of technical progress, even if capital does not grow faster than labour, per capita income may not decrease provided technical progress is high enough to offset the effect, $\varepsilon\dot{L}$.

Recursive and Interdependent Models

Before we say something about short-term dynamic analysis, let a reference be made to a controversy concerning whether economic theory is best expressed as a system of recursive simultaneous equations or as a system of interdependent simultaneous equations. An example of a recursive model is given below:

$$Y_t = C_t + I_t$$
$$C_t = a.Y_{t-1} = \text{(in general) } f(Y_{t-1})$$
$$I_t = b.Y_{t-1} = \text{(in general) } F(Y_{t-1})$$

An interdependent model is as follows:

$$Y_t = C_t + I_t$$
$$C_t = a.Y_t = \text{(in general) } f(Y_t)$$
$$I_t = b.Y_{t-1} = \text{(in general) } F(Y_{t-1})$$

In the first case, both consumption (C) and investment (I) are determined by last year's income; and these, in turn, determine this year's

income. But in the second case, consumption and income for the current year are interdependent. If we use arrow diagrams we can depict the two models as in Diagram 7.1.

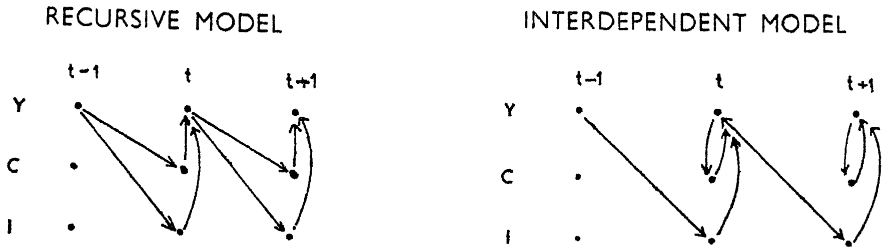


DIAGRAM 7.1

In the first case, all the arrows within a period run in one direction, but in the second case the arrows within a period run in opposite directions at least once.

Since human reactions take time, interdependency can be ruled out under perfect competition. Where two individuals consult each other before taking decisions, e.g., a sales manager and a purchase manager, they belong (or, may be taken to belong) to the same behaving unit: so, their individual importance is insignificant. Similarly, in the absence of competition, where parties consult each other, we may well treat all concerned parties as one behaving unit: alternatively, we may write a model to embrace all negotiation-sequence and this will be naturally recursive. Each recursive equation allows for a causal interpretation. Besides, the recursive character fits in with the line of thought that *expectations* lead to *plans*; plans lead to *actions*; actions lead to taking stock of realisations, i.e., to *accounting*: realisations lead to *new* expectations, and the causal cycle is repeated. Incidentally, when it is said that dynamic analysis is inevitably related to expectations, it must be borne in mind that expectation is a function of past realised values.

If Y', C' and I' are used to indicate expected or ex-ante income, consumption and investment, we may write a model as follows:

1. Expected income this year is a function of last year's income:

$$Y_t' = a_0 + a_1.Y_{t-1}$$

2. Consumption is planned on the basis of expected income:

$$C_t' = c_0 + c_1.Y_t'$$

3. Planned investment depends on last year's income:

$$I_t' = b_0 + b_1.Y_{t-1}$$

If we assume that consumers and entrepreneurs always succeed in realising their plans, then C_t equals C_t' and I_t equals I_t'. Hence C_t and I_t become known and so

$$Y_t = C_t + I_t$$

This Y_t will now determine the expected income for the next year.

If we eliminate the ex-ante values, the model becomes:

$$Y_t = C_t + I_t$$
$$I_t = b_0 + b_1 Y_{t-1}$$
$$C_t = c_0 + c_1.(a_0 + a_1 Y_{t-1}) = d_0 + d_1.Y_{t-1}$$

where $d_0 = c_0 + a_0.c_1$ and $d_1 = a_1.c_1$. Here we see that the student of statistics will not be able to tell us the value of a's and c's beyond d's.

If, however, we further assume that the difference between actual and expected income is random or stochastic, then we can write the third equation about consumption as follows:

$$C_t = c_0 + c_1 (Y_t - e_t) = c_0 + c_1 Y_t + e_t'$$

where e_t is the error-term and $e_t' = c_1 e_t$.

Although the basic equations were recursive, the first derived set of equations retain the recursive nature but the second set of equations takes an interdependent shape. These equations can yet be estimated: only their worth may be questioned if the economy is a developing economy suffering from inflation, for then the deviation between ex-ante income and realised income cannot be termed random or stochastic in the statistical sense.

Short-term Dynamic Analysis

The cobweb model as also other simple multiplier-accelerator models in economics may be said to be related to short-term dynamic analysis. These throw light on how fluctuations take place in an economy.

To give an example of a short-term dynamic analysis for a producer, let us assume that production over time is proportional to excess of marginal revenue (MR) over marginal cost (MC):

$$\frac{dx}{dt} = a\,(MR - MC) = \text{(say) } a\left(\frac{dT\,(x)}{dx} - \frac{dC\,(x)}{dx}\right) = a\,(T' - C') \text{ (in brief)}$$

where $T(x)$ and $C(x)$ are total revenue and total cost functions.

If we want to study the behaviour of production round a value, x_0, we may expand the T and C functions as Taylor series and, as a first approximation, we may write:

$$\frac{dx}{dt} = a\{T'(x_0) + (x - x_0)\,T''(x_0) - C'(x_0) - (x - x_0).\,C''(x_0)\}$$
$$= a\{T'(x_0) - C'(x_0)\} + a\{T''(x_0) - C''(x_0)\}\,(x - x_0)$$

If x_0 be an equilibrium, or near-equilibrium, point, then $T'(x) - C'(x)$ will be nearly zero and may be neglected so that

$$\frac{1}{x - x_0} \frac{dx}{dt} = a\,[T''\,(x_0) - C''\,(x_0)] = \text{(say) } az$$

Then, the solution is

$$x - x_0 = A.e^{azt}$$

where A is a constant. If the marginal cost curve intersects the marginal revenue curve from below, z will be negative and x will cyclically tend to x_0 in due course. The coefficient a may be conceived to hide within it many effects, institutional, technical and even that relating to subjective sensitivity of producers.

If we introduce here planned purchases by consumers and price policy on the part of the seller, we may write:

1. Consumers try to adjust purchase to planned purchase (x^*):

$$\frac{dx}{dt} = m\,(x^* - x)$$

where x^* is planned purchase, and x is the actual purchase.

2. The producer tends to raise price if sales (x) exceed production (x_p):

$$\frac{dp}{dt} = n\,(x - x_p)$$

3. The producer raises output if marginal revenue exceeds marginal cost:

$$\frac{dx_p}{dt} = a\,[T'\,(x) - C'\,(x)]$$

For equilibrium the right-hand-side expressions of the three equations should be equated to zero.

These can be generalised for two or more producers.

It may be added that in connection with consumption functions, a number of forms have been studied econometrically. We have mentioned above:

$$\frac{dx}{dt} = m\,(x^* - x)$$

Another form, called the *DEF* model, is in terms of consumption ratio (r_t): and ratio (z_t) of income (Y_{t-1}) to highest income (Y_{max}) reached in the past:

$$r_t = C_t / Y_{t-1}$$
$$z_t = Y_{t-1} / Y_{max}$$

The actual consumption function is written as follows:

$$r_t^* = u + vz_t$$
$$r_t - r_{t-1} = w\,(r_t^* - r_{t-1})$$

where r^* is the desired consumption-ratio. From these, on eliminating r's and z, we get a relationship of the form—

$$\frac{C_t}{Y_{t-1}} = \alpha + \beta \frac{Y_{t-1}}{Y_{max}} + \gamma \frac{C_{t-1}}{Y_{t-2}}$$

which has been used for estimation purposes.

There are other consumption functions used and estimated but it is not easy—in general, it is difficult—to produce the economic theory that can justify them.

Short-term dynamic analysis is widely applied to the study of economic phenomena. Here, one finds a study of dynamic stability of a potato market; there, one reads of dynamics of wages, capital costs and employment in manufacturing. Studies are also being made of the nature of expectations of manufacturers about industrial sales and of dealers about retail sales. Models have been formulated to explain businessman's behaviour. In the post-Second-World-War period businessmen have been found to have become more optimistic; and their expectations have possibly important stabilising properties in cyclical fluctuations of the capital goods sector: and in the case of sales of non-durable goods the anticipated trend in sales has been found to be adjusted to the extent of about 20 per cent of the discrepancy between the anticipated and actual sales of the recent past.

There are, no doubt, economic models behind such studies as are referred to above, but the models are not based on economic theory: rather these models are likely to be the basis for new economic theory.

The various trade cycle dynamic models are also examples of short-term dynamic analysis. We may illustrate them with just one example although it may mean some repetition:

$$Y_t = C_t + I_t + A_t$$

where Y, C and I have their usual meaning and A stands for autonomous investment, which may (or may not) be a function of time, i.e. which may in general change with time in some regular manner.

$$C_t = c_1.Y_{t-1} + c_2.Y_{t-2}$$
$$I_t = b\,(Y_{t-1} - Y_{t-2})$$

so that, on eliminating C_t and I_t we get

$$\begin{aligned} Y_t &= (c_1 + c_2)\,Y_{t-1} + (b - c_2)\,(Y_{t-1} - Y_{t-2}) + A_t \\ \text{(say)} &= cY_{t-1} + w\,(Y_{t-1} - Y_{t-2}) + A_t \\ &= (c + w)\,Y_{t-1} - wY_{t-2} + A_t \end{aligned}$$

This is a second order difference equation. Assuming the autonomous investment to be constant for the time being, the time path of Y_t can be studied.

If $c + w = 2$, $w = -3$ and $A = 8$, and $Y_0 = 6$ and $Y_1 = 2$, we find, as Baumol did in his *Economic Dynamics* (p. 171), that

$$Y_t = 3.3^t + 5(-1)^t - 2$$

Or, as Allen has shown in his *Mathematical Economics* (p. 211), we may characterise the time-path of Y for different values of w:

Condition regarding w	*Nature of time-path of* Y
1. w is less than z_1	Non-oscillatory and damped
2. w lies between z_1 and 1	Oscillatory and damped
3. w lies between 1 and z_2	Oscillatory and explosive
4. w is greater than z_2	Non-oscillatory and explosive

Note: $z_1 = (1 - \sqrt{1-c})^2$ and $z_2 = (1 + \sqrt{1-c})^2$

We have in economic literature knowledge available on dynamic analysis of flow models, stock models and 'stock and flow' models for one, two and more commodities. Samuelson discussed stock and flow theories in Part II of his *Foundations of Economic Analysis* and Klein (L.R.) wrote about 'Stock and Flow Analysis in Economics' in *Econometrica* (July, 1950). We also have a treatment of dynamic macro-models in Baumol's *Economic Dynamics*, Allen's *Mathematical Economics*, Goodwin's contribution to *Econometrica* (January, 1951), and in the writings of Hicks, Samuelson and others.

For purposes of understanding the real world, economics and analyses go a very little way. The ratio of enlightenment to existing complexity is so small that notwithstanding dynamic models and dynamic analysis, 'intuition'—considered by many as a substitute of dubious value—'enlightened intuition' or 'wantless intuition' are ultimately essential. Dynamic and static analysis, whose extent is being increased, provides desirable assistance.

CHAPTER 8

Social Accounting

INTRODUCTION

SOCIAL accounting is a method of studying 'the *structure* of the body economic' while a study of the *working* of the body economic (or, the economic system) is called the theory of value. Social accounting and the theory of value are like anatomy and physiology in medical science: while anatomy studies the structure of the body, its organs and their relations, physiology studies the actual working of the body. Similarly, social accounting is a technique of presenting information about the nature of the economy of a society (institutions, state or nation) with a view not merely to get an idea of its prosperity, past or present, but also to get guidelines for collective (*or* state) policy to influence (*or* regulate) the economy. It is concerned with classification and presentation of activities of human beings and human institutions so as to provide us with a new way of understanding the operations of an economy. Really, social accounting came into prominence after the great depression of the 1930's, and particularly during and after the Second World War. *Now*, it is taken to be essential for planned economic development, and some interpret it to include not only classification but also application of the compiled information (*i*) to investigate how an economy works, and (*ii*) to indicate future lines of action. Social accounting statistics and social accounting models of an economy are both now included within the scope of social accounting. A certain section of students of economics is now specialising in this field and we may well have one day, besides

economic advisers, social accountants, social accounting experts and social accounting advisers.

DOUBLE-ENTRY ACCOUNTS

As a term, social accounting was coined on the analogy of private accounting and is usually associated with the double-entry method of private accounting. Any private producer or trader maintains his accounts these days on the double-entry accounting basis. To give an illustration, let a trader have only two customers, *X* and *Y* to whom he sold goods worth Rs 400 and Rs 300 and from whom he received Rs 295 and Rs 280 during a given period. We would find the following accounts in his account books:

	X'S ACCOUNT		
To Stock	Rs. 400	By Cash	Rs 295
	Y'S ACCOUNT		
To Stock	Rs 300	By Cash	Rs 280
	CASH ACCOUNT		
To X	Rs 295	By —	Nil
To Y	Rs 280		
	STOCK ACCOUNT		
To —	Nil	By X	Rs 400
		By Y	Rs 300
	Rs 1,275		Rs 1,275

Here for each transaction there are *two* entries, one on each side. If we add up the entries on the two sides separately for all the four accounts, the two totals will be the same. The trader does test the correctness of his accounts in this way, by drawing up a *trial balance.*

Again, the transactions are on an *accrual basis,* i.e., a transaction is registered even though it is not immediately paid for. In certain respects, the accrual entries are made at the time of closing the annual accounts. Thus accounts for 'interest due', and 'depreciation allocation on capital goods used' get their entries only at the year-end. To take an example, let the trader take a loan of Rs 10,000 from *Z* at 6% per annum and purchase a machine at the beginning of the year: let him pay Rs 400 on interest account during the year. Also, let depreciation be reckoned at 5% yearly. Then we shall find in his account books the following entries, the left- hand entries being made at the close of the year:

	INTEREST ACCOUNT		
To *Z* (for interest)	Rs 600	By Profit & Loss A/c	Rs 600
	DEPRECIATION ACCOUNT		
To Stock	Rs 500	By Profit & Loss A/c	Rs 500
	Z'S ACCOUNT		
To Cash	Rs 400	By Cash	Rs 10,000
		By Interest	Rs 600

The Profit and Loss Account is not a regular account in the account books. It is drawn up once a year and is closed by transfer to the *balance sheet* which is a statement showing the various assets and liabilities of the trader. The interest and depreciation entries therefore do not have corresponding second entries in the accounts of the trader. Yet the system of accounting as a whole is said to be on the double-entry basis. The point to be noted for the future is that some entries are on a payment basis and some on an accrual basis. Similarly, under social accounting, everything is not on an accrual basis, but we shall come to this again.

SOCIAL ACCOUNTS

As in private accounting, we may draw up double-entry social accounts for an economy. We give in the accompanying table examples of such accounts for the divisions: (*a*) firms, (*b*) households, (*c*) governments and (*d*) stocks (or capital) of a closed economy:

Receipts		*Payments*	
		FIRM'S ACCOUNT	
Firms	400	Firms	400
Households (for goods)	230	Households (for labour)	190
Government (for goods)	37	Government (indirect tax)	70
Government (subsidy)	2	Capital	12
Capital	3		
	672		672
		HOUSEHOLD'S ACCOUNT	
Firms	190	Firms (for goods)	230
Households (for service)	10	Households (service)	10
Households (for gifts)	3	Households (gifts)	3
Government (transfer payment)	60	Government (direct tax)	29
Capital (for interest)	11	Capital (loans given)	2
	274		274
		GOVERNMENT ACCOUNTS	
Firms (indirect tax)	70	Firms (for goods)	37
Households (direct tax)	29	Firms (subsidy)	2
		Households (transfer payments)	60
	99		99
		CAPITAL ACCOUNT	
Firms	12	Firms	3
Households	2	Households	11
	14		14

Each entry in these four accounts occurs twice, once on the left-side as a receipt from some head and again on the right-side as an expenditure item. Besides, each of the four accounts is itself balanced: here no account shows expenditure exceeding or falling short of income.

SOCIAL ACCOUNTS TABLE

The information contained in both private accounts and social accounts can be simplified by expressing it in the form of a 2-way table as shown below:

PRIVATE ACCOUNTS TABLE

Received by	*Payment by*					*Total*
	X	*Y*	*Z*	*Cash*	*Stocks*	
X	—	—	—	—	400	400
Y	—	—	—	—	300	300
Z	—	—	—	400	—	400
Cash	295	280	10,000	—	—	10,575
Stocks	—	—	—	10,000	—	10,000
TOTAL	295	280	10,000	10,400	700	

SOCIAL ACCOUNTS TABLE

Received by	*Payment by*				*Total*
	Firms	*Households*	*Government*	*Capital*	
Firms	400	230	37+2	3	672
Households	190	10+3	60	11	274
Government	70	29	—	—	99
Capital	12	2	—	—	14
TOTAL	672	274	99	14	

In the private accounts table all transactions are not paid for at once and hence the total receipts and total payments for each *sector* (*X*, *Y*, *Z* Cash, Stocks) are not equal. This is not so in the case of the social accounts table, wherein we *seem* to start from a clean slate every time and payments are made for value received during the year: each sector is therefore seen to equate both sides of its accounts. There is no 'bringing forward' and 'carrying over' of assets and liabilities and there are no year-end adjustments.

Many other classifications can be adopted. Also, the scope of a class and the number of classes may be changed. Thus we give below three other social accounts table, two of them with illustrative figures:

I

Received by	*Payment by*			*Total*
	Firms	*Households*	*Stocks*	
Firms	50	20	30	100
Households	30	5	—	35
Stocks	20	10	—	30
TOTAL	100	35	30	

II

Received by	Payment by					Total
	Firms	*H.H.*	*R.W.*	*Govt.*	*Capital*	
Firms	...	...	...	...	...	...
Households (H.H.)	...	...	...	...	...	...
Rest of world (R.W.)	...	...	...	...	...	...
Government (Govt.)	...	...	...	...	...	...
Capital	...	...	...	...	...	...
TOTAL	...	...	...	...	...	...

III

Received by	Payment by				Total
	Production	*Consumption*	*Accumulation*	*R of W*	
Production	—	11,324	1,431	870	13,625
Consumption	11,973	—	—	592	12,565
Accumulation	593	1,041	—	549	2,183
Rest of world	1,059	200	752	—	2,011
TOTAL	13,625	12,565	2,183	2,011	

FLOW DIAGRAMS

Social accounting information can be shown to an elementary extent by means of diagrams also. In such diagrams arrows indicate the directions in which money-values flow from and to the heads (*or* sectors) shown in small rectangles (Diagram 8.1).

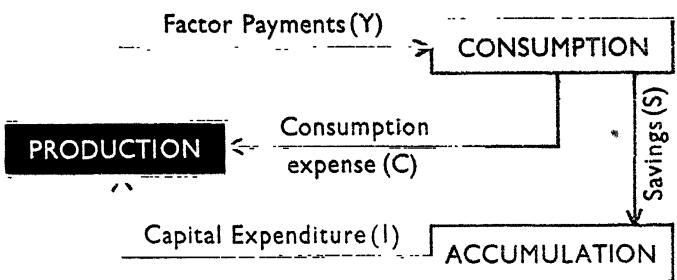


DIAGRAM 8.1

NOTE: Using the symbols indicated within brackets in the diagram for the flows we may, for equilibrium at each rectangle, write out the following model-relationships:

Production end: $Y = C + I$

Consumption end: $Y = C + S$

Accumulation end: $S = I$

If we were to take account of transactions with the rest of the world also, a flow-diagram could be drawn up as showing in Diagram 8.2.

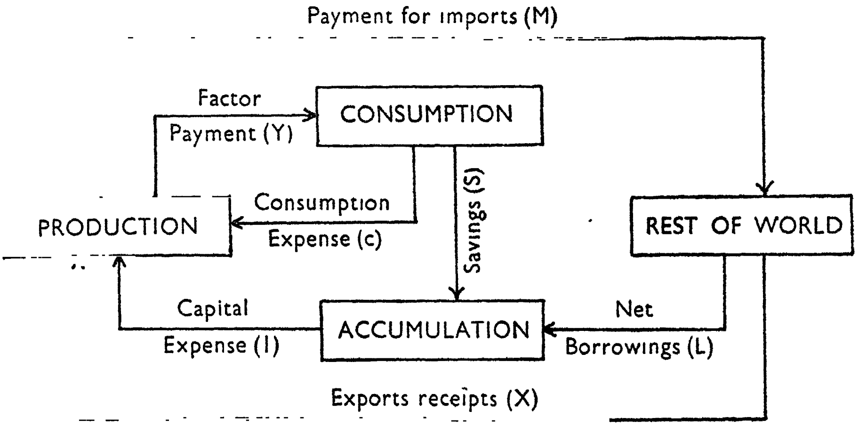


DIAGRAM 8.2

NOTE: Using the symbols as before we could write:

Production end: $Y + M = C + I + X$

Consumption end: $Y = C + S$

Accumulation end: $I = S + L$

Rest of world end: $M = X + L$

and we may also put $Y = C + I - L$

We could make the second flow-diagram more detailed by adding more flow lines as shown in Diagram 8.3.

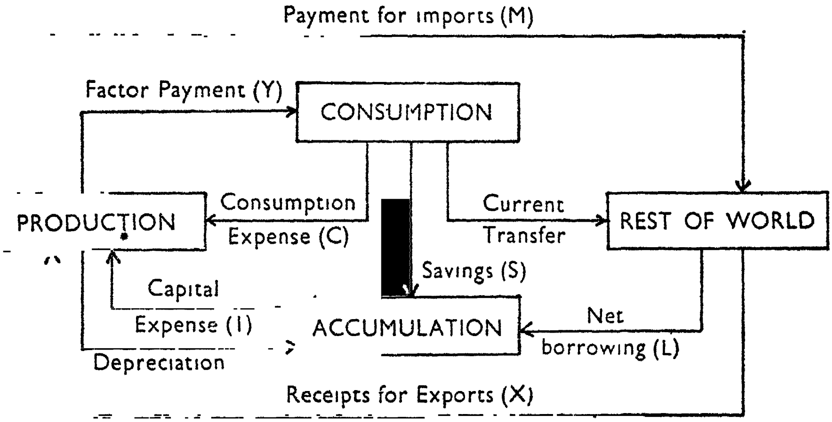


DIAGRAM 8.3

CLASSIFICATION IN SOCIAL ACCOUNTS TABLE

Classification in a social accounts table may be (usually is) practised so as to take account of

(1) forms of economic activity such as 'production', 'consumption' and 'addition to wealth';

(2) institutional (or sectoral) subdivisions of the economy such as 'firms', 'households', 'government' and 'rest of world'; and

(3) transactions such as 'sales and purchase', and 'current transfer'. Thus, when we use the classes production (*P*), consumption (*C*), and accumulation (*A*) in a flow-diagram (and a social accounts table can be easily drawn up as shown alongside) we consider and take into account forms of economic activity. When we draw up the table by firms, households, rest of world and government also, we take account of the institutions as well. Incidentally, it is usual to consider households and government to constitute together the activity group called 'consumption'. Similarly, we can draw up a table for transactions.

	P	*C*	*A*	*Total*
P	...	...	...	...
C	...	...	...	...
A	...	...	...	...
TOTAL	...	...	...	...

In general, a detailed social accounting table (*usually* called a social accounts *matrix*) may be somewhat as shown in Diagram 8.4.

The subdivisions of the production sector are shown to be 10 in this table: they can be more than 10. However, usually, the last production sub-sector is 'other production and trade' and covers transport, communication, distributive trades, insurance, banking, finance and other services: it accounts only for their running costs and for their services and *not* for their gross purchases and sales.

The consumption sector is vertically divided into 'business and households' and 'government' sectors.

The 'government-sector' column usually includes payments made for such government sections as public administration, defence, public health, education, 'domestic services to households' and 'services to non-profit making bodies'. Any government industrial undertaking finds a place in the relevant production sector.

The 'accumulation' sector is sometimes labelled 'capital transactions'.

The 'external' sector refers to rest of world (ROW).

When we come to the rows which indicate how the different sectors send out their goods and services, i.e., how they receive their payments, it is usual to divide the 'business and household' sector into (*a*) wages and salaries, (*b*) profit, interest and rent, (*c*) interest earned on public debt, and (*d*) other receipts through current transfers.

Similarly, the row for the 'accumulation' sector is subdivided into: (*a*) Depreciation, (*b*) Savings, (*c*) Capital Transfers (net).

The 'external' sector row is subdivided into (*a*) goods and services imported, (*b*) current transfers (net), including factor earnings by nationals from abroad, and (*c*) lending (net) made to foreign countries.

Incidentally, it may be mentioned that payment by consumers to

		Production subdivided into 10 sectors of industries										Consumption		Accumu-lation	External (ROW)	Total
		1	2	3	4	5	6	7	8	9	10	Busi-ness & H H	Govt			
Production (10 divisions)	1															
	2															
	3															
	4															
	5															
	6															
	7															
	8															
	9															
	10															
B & H H	Wages & Salary															
	Pft, Int, Rent															
	Public Debt Interest															
	Other current transfers															
Govt	(Tax less subsidy)															
Accumn	Depreciation															
	Savings															
	Capital Transfer															
R O W	Goods and Services															
	Current Transfer															
	Net lending															

Note: B + H H = Business and Household
Govt = Government
Accumn = Accumulation
ROW = Rest of World

DIAGRAM 8.4

'production' does not distinguish between durable consumer goods and perishable consumer goods: the durable consumer goods are taken as written off the moment they reach the consumers.

SOCIAL ACCOUNTS AND INPUT-OUTPUT MATRIX

An input-output matrix may be conceived as a table showing flows to and from producers in the economy, divided according to different branches of production. But statistical data are generally available institution-wise. Thus, though a firm may be manufacturing both fertilisers and plastics, it may not be feasible to separate (or get separate) data for fertilisers (to be included under agriculture). This heterogeneity affects the assumption of stability of technical coefficients. However, in

practice, if the institution (say, a firm) has separate establishments—say, one for oils and the other for toys—and if separate data are available, the establishments are classified under separate appropriate commodity production heads. Where a firm produces a subsidiary product, it is logical to carry the details of the subsidiary product to its relevant production group. This is practised as far as possible though a number of assumptions, as far as possible reasonable, have to be made.

In the case of byproducts, however, the convention is to show them as negative input in the main industry.

With these observations, one may say that if we retain the rows and columns corresponding to production but consolidate all other columns and rows we can get the input-output matrix from a social accounts table. But this cannot serve as a good forecasting instrument as all durable goods and raw materials are shown as used up as soon as purchased. Besides, the 'accumulation' entry in each row is not only in respect of stocks of finished industrial product but also in respect of different raw materials.

Another problem is in regard to prices at which elements in a row are evaluated. To some sectors goods are sold at wholesale prices; to the final consumers they are sold at retail prices. Hence the money valuations of elements in a row are not likely to be at the same prices. A *better* course can be (though it will not be realistic) to show all elements at producer's price and to put the distribution cost borne by final consumers as that paid to the 'other production and trade' sector.

Yet another point concerns imports. The 'external' sector row for goods and services should show under a production sector only those imports which are complementary to the production sector: all other imports shuold be classified elsewhere, probably under 'business and household'.

There is yet another aspect. Suppose we import chemicals for an industry which purchases the same chemicals indigenously also. Then, unless the imported chemicals are shown under the said industry (and not under the appropriate chemical industry), it will not be easy to estimate the effect of changes in that industry on the consumption of that chemical. A solution is to add such import to local supply of chemicals to the industry and to deduct (if possible) the amount (i.e., to show it as negative export) out of the exports of chemicals.

Social Accounts and National Income Accounts

Not infrequently, economic accounting, national income accounting and social accounting are treated as synonyms. It would be better if, instead of national income accounting, the term 'national accounting'

is used. For, national income 'accounts' are an estimate of the increase in national wealth in a year. Or, the 'accounts' may be said to be gross production expenditure (*or* gross production income) less depreciation. It will be the consolidated total for the production sector, whether for the column or the row. Thus national income accounts may be represented (*i*) either as a row entry, or (*ii*) as a column entry, less (of course) depreciation.

But this raises a few points for consideration. Firstly, the 'external column' shows payments for exports and includes payments arising on account of (*i*) returns on property held abroad, and (*ii*) factor income received from abroad. These are justifiably included in 'national income' and hence should be transferred to the 'firms' column.

Similarly, national income (or expenditure) is taken to include (*i*) not only the element due to government production activities such as electric undertakings and steel mills, but also (*ii*) government expenditure on services by individuals as administrators, soldiers and policemen. Kuznets once raised the question whether these should not be taken to be prerequisites of the social framework in which economic activities arise. If the answer given is 'yes', then we cannot include these service-costs in national income. Otherwise, we will have to shift these to the production sector for purposes of national income accounts: in practice, this is what is usually done, perhaps because it is calculated to give a "better view of the total resources" that are available as a flow. Even so, a controversy is bound to remain in regard to what part of government expenses on durable goods and services, showing effects over a number of years, should be excluded.

Also, the firms column includes, against the government row, both direct and indirect taxes paid minus subsidy (if any) received from government. Since indirect taxes should not enter the column for purposes of national income accounts, these have to be subtracted.

Even so, we shall get national income *at factor cost*, i.e., at cost of production. This will differ from *gross national expenditure* which is incurred at *market prices* and which includes:

(*i*) Household purchase of consumer goods
(*ii*) Government purchase of current goods and services
(*iii*) Gross domestic investment
(*iv*) Net investment abroad

The gross national expenditure is also called *national income at market prices* and includes indirect taxes.

Incidentally, the argument for conceiving national income at factor cost is that indirect taxes may be different over time and regions and these would make it difficult to secure comparability over time and region; yet it is not free from (secondary) effects due to changes in direct taxation.

Applications in Economics

The development and use of social accounting in economics has been the result of (*a*) a shift of interest from microeconomics to macroeconomics, (*b*) studies to understand macroeconomic repercussions, and (*c*) the use of statistical methods to build up macroeconomic models in practice.

At the national level, 'state policies of full employment', and national budgeting and the assessment of economic welfare are made easy, thanks to social accounts. Also, in this age of central planning, we need to know the economic position of different economic classes and economic regions.

Today, nations are making economic models, either short-term forecasting models or long-term planning models, which vary in content and form: but, invariably, they make use of a social accounting framework. These models do enable us to understand what happened in the past.

Besides, social accounting data help 'the partial users' both: (*i*) the economist who is interested in products, inputs, outputs and employment, and (*ii*) the student investigating the financial side, viz., income, savings and flow of funds.

At the international level, and at times even for drawing lessons at the national level, inter-country comparisons are made for purposes of common defence and group-economic-decisions. The creation of such bodies as the U.N., the I.M.F. the E.E.C., and E.F.T.A. has increased the need to assess national income and prepare social accounts.

Limitations

Two limitations deserve mention. One continues to exist, the other is being remedied.

Though not so much in natural sciences, where controlled experiments are more feasible and where individuals may be ignored, in economics it is difficult (*i*) to foresee when generalisation leads to a bad decision, and (*ii*) to ignore the individuals and lose sight of the particular. Thus a generalisation that "people in general did not die of starvation in east India" may not be acceptable if it hid a fact that "two (or five) percent did die of starvation". While such a (2 or 5%) variation may be tolerated in the manufacture of goods, it is not likely to be tolerated where human lives are concerned.

The other limitation refers to the lack of an account (like that shown by a balance sheet) indicating the assets and liabilities of the economy.

Social accounts do present us with a picture of the flows during a period (usually a year) and enable us to understand the relative importance of the various sectors as sources and absorbers of flows: they do help the decision maker and economic planner to take a better decision on what they can tap, and how. But the year-end balance sheet of the stock position of the various sectors is lacking. The latter is no doubt necessary in order (*i*) to judge the relative importance of the changes as a result of the flows, and (*ii*) to plan for the extent of changes to be sought in the future. There are plans and attempts to take a census of national wealth. Even so, there may be left out, both in the national accounts of flow and new national accounts (or balance sheets) of assets and liabilities, much of what cannot be (or, what is not usually) measured in money.

Present Trends

At the United Nations level there is being evolved a revised system of national accounts with the following eight major classes (including sub-classes) for a two-way table:

(1) Opening balance sheet
(2) Current accounts:
  (*i*) Production
  (*ii*) Income and outlay
(3) Capital accounts:
  (*i*) Capital expenditure
  (*ii*) Capital finance
(4) Rest of world account
(5) Revaluations
(6) Closing balance sheet

It is planned to satisfy at least five requirements.

(*a*) Provision of a complete set of arithmetical and accounting checks.
(*b*) Reflection, within the frame, of important economic and accounting distinctions; such as home and external, current and capital, real and financial, and stocks and flows.
(*c*) Indication of compromises between what is analytically desirable and what is statistically feasible.
(*d*) Relation of various classifications needed to connect the different parts of an economy.
(*e*) Avoidance of unnecessary subdivisions of a category.

It is not opportune to go into further details, expositional or critical, in this regard. The revised 'system of national accounts' has been published and an excellent appraisal by Prof. R. Stone is available in the first issue of *The Review of Income and Wealth*.

CHAPTER 9

Econometrics

PURE *vs* APPLIED ECONOMICS

THE TASK of a scientist is to discover relationships between the phenomena that lie within the universe of his study; and the relationship that most concerns him is the temporal relationship. His explorations in the field of the phenomenal world relate to events that have location in space and time. Every phenomenon he studies is subject to space and time limitations. The relationships that he discovers must, therefore, be spatial or temporal. But, as observed above, it is the temporal relationship that is most essential for his purpose and it is from knowledge of such relationships that he can hope to make predictions about the future. A man lives in the present but the immutable part of his being is beyond the limitation of time and it exists, spread as it were, over the limitless path of time. The instinctive longing to know the future is, therefore, but natural. And this explains his concern with the knowledge of temporal relationships between the phenomena that constitute the environment within which he physically and psychologically dwells.

The discovery of temporal relationships requires a careful observation of all relevant entities and events. Since such a relationship is most essentially causal in nature, a scientist has, in other words, to find the cause and effect of every event. There are various ways in which he can proceed to make such an investigation; the laws of logic and the various methods of discovering causal relationship come to his aid. Of these methods the one that makes perhaps the least demand on his patience

and the material resources at his command is the method of concomitant variation. But for the application of this method he must be able to subject his material to the process of measurement. The word measurement has a wide connotation in science; there are degrees and kinds of measurement, each suitable for a particular purpose. For instance, there is measurement that is correct up to monotonic transformation, one which supplies knowledge of the direction of change. We infer causal relationships or temporal relationships from such knowledge regarding the direction of change of the phenomena under investigation. When demand increases price rises; when supply increases price falls. These are the relationships that involve measurement up to monotonic transformation.

MEASUREMENT IN APPLIED ECONOMICS

Economic theory, pure theory, deals in such relationships. It is able to supply the economist with causal relationships, the extent of which is limited by measurement that is confined to the direction of change in the entities he handles. To what extent the demand will increase when price falls is a piece of information that pure theory is incapable of supplying. For, pure theory, in its purity, is based on the most fundamental knowledge in regard to human behaviour. And such behaviour, by its common pattern, reveals a causal or temporal relationship that is confined to the direction of change only.

Those who seek no more knowledge need nothing besides pure economic theory. A theorist lives in a region where for his subsistence no further measurement is needed. He trades, as it were, in broad generalities and does not need to particularise his knowledge as an applied economist does. His predictions are in general terms and for such predictions he needs the type of knowledge that his observation of human behaviour is able to supply him. But an applied economist needs more detailed and particularised knowledge. He is interested in making predictions that have more precise implications. He is willing to sacrifice generality in his conclusions for the sake of knowledge that has more localised application. His findings are comparatively more phenomenalised than those of a pure theorist. While the phenomena a theorist finds necessary to measure are those that are defined by their direction, those that an applied economist has to measure have additional attributes. He has to establish relationships not merely between magnitudes but also between differences of magnitudes. He has, as it were, to trade in differences, in relatives rather than absolutes, in rates of change rather, than changes themselves.

While pure theory is capable of establishing a relationship between

change of demand and change of price, for example, an applied economist needs tools that will enable him to establish relationships between the rate of change of demand and rate of change of price. While the theorist handles the first-order phenomena the applied economist handles higher-order phenomena. Pure theory that relies on fundamental knowledge of human behaviour is unable to provide a technique competent to measure the higher order phenomena. But the pure economist has no laments for this. He dwells in regions where the necessity for such a technique does not exist. There are hardly any constraints on his imagination and if his explorations still have relevance to the type of beings that an applied economist studies that is because, while his mind takes flights of imagination, his feet are yet planted on the material world in which he physically exists. But there is a sharp difference between the objectives of a pure scientist and those of an applied one. The latter has to have his propositions more phenomenalised, more particularised, more firmly localised. His search, therefore, for techniques that will enable him to measure higher-order phenomena is understandable. These are necessary for his purpose and when he uses these tools, applying them to the unsatisfying knowledge of pure theory, he creates a new discipline, the captivating discipline of econometrics.

BIRTH OF ECONOMETRICS

This is how econometrics takes birth; it is born of the marriage of pure economic theory with statistics and mathematics. Now let us talk in less abstract language. The pure economist says, for instance, that demand is a function of price. This functional relationship is most generally expressed by an equation of the form, $x = f(p)$, where x stands for the quantity of a commodity demanded and p for the price. But this means very little for the applied economist and he wants to give this relationship a more precise shape. The pure economist knows that the above equation plotted as a graph must be a negatively inclined curve. He knows this but cannot say what shape it will have. He knows, in other words, that $f'(p)$ is negative but cannot say what is the sign of $f''(p)$. This is what we meant by saying that the pure economist is unable to measure second-order phenomena. Nor does he know, even if he knew the exact shape of the curve, at what height it is located for any particular market. The applied economist must know whether the demand equation is linear, quadratic or cubic. He must know the values of the parameters and the values of the coefficients of the variables, demand and price.

When the applied economist knows the precise nature of the equation of the demand curve he is able to draw it and give it a location and a shape. He can then set it against a similar supply curve and find out

the price. He is interested in making particularised predictions and for this he needs knowledge of the kind described above. The equation $x = f(p)$ can be made more precise, for instance, by giving it either a linear, a quadratic or a cubic shape. We thus have a large number of equations to choose from. A linear equation can be of the form $p = ax+b$. A quadratic equation can be of the form $p = ax^2 + bx + c$. A cubic equation can be of the form $p = ax^3 + bx^2 + cx + d$. These give us some idea of the general shape of the demand curve. But for precise knowledge of the shape as also for the location of the curve we need to know the numerical values of the coefficients: we must know the values of the parameters.

To obtain knowledge of this type the economist has to enter the market place and make a search there with the help of statistics and mathematics. He has to plot the data relating to price and demand and then fit a curve into them. There are methods to be employed for this purpose.

Model Building

An econometrician has to proceed by stages. He has to have a skeleton of bare bones before he can begin to stuff it with flesh and blood. The skeleton is supplied by pure theory and consists of a model expressed in general terms. A model in economics consists of a set of equations that express relationships between entities that are relevant to the problem under investigation. These entities are called variables. In economics, the variable entities that we often need to use are income, savings, investment, supply of money, capital-stock and so on. These variables are very relevant to the study of macroeconomic problems. Equations indicating relationships between these entities, when taken together, make up a model. A simple microeconomic model relating to a particular commodity would consist of demand and supply equations such as

$$d = ap + b \qquad \text{and} \qquad s = \alpha p + \beta$$

These are linear equations meaning that, when traced, they turn out to be straight lines. Here, in this model, the entities used are price, quantity demanded and quantity supplied. These equations considered together give rise to the concept of a model. The object is to find the price and the quantities demanded and supplied when the system attains equilibrium position. These three entities are called variables, to determine the values of which the equations have to be simultaneously solved.

To repeat, then, the work of an econometrician starts with the building of a general model. The model consists of equations giving relationships between relevant entities called variables. We have taken the simplest possible example of a microeconomic model. It is simple because we have taken linear equations and it is the simplest because there are only

systematic variables in the equations. The quantities demanded and supplied are variables which behave in an orderly manner. We call their behaviour orderly only because some dependable guess can be made about the manner in which they vary. Sometimes we say that these (systematic) variables show a system in their behaviour and because of this their behaviour can be described as predictable. In a more complicated or a less simple model there are, at times, variables which are called random. As the word suggests, such variables behave in an unpredictable, disorderly fashion. As a matter of fact there is no disorder in nature and where the behaviour of an entity appears to show no system it is because our knowledge is imperfect. Be that what it may, where we do not know how an entity behaves or where our knowledge is not sufficiently perfect for our purpose, it becomes a random variable. In the above equations, for example, the quantity demanded is shown to depend on price and a given constant b. It can also depend on fashion but how fashion would change from time to time we do not know, nor do we know what precisely would be the nature of the relationship between it and demand. To take account of such an unsystematic relationship between demand and fashion we can add to the equation, say, u, and call it a random variable.

Particularisation of a Model

A model, as described above, consists of a number of equations that indicate relationships between significant entities. These relations are of a very general nature such as those between (*i*) the direction of change in price and the direction of change in demand, and (*ii*) the direction of change in income and the direction of change in consumption. These relationships are obtained from an intelligent observation of human behaviour. It is here that pure economic theory plays its part. Since all relationships are quantitative in the final analysis they are based on measurement of entities. And, as explained earlier, pure theory is concerned with the measurement of first-order entities. This is why the model supplied to us by pure theory are in terms of algebraical equations. All quantitative relationships in such equations are represented by symbols such as a, b, c, x, y, s, d and u.

An applied economist is not satisfied with the use of such symbols. He needs more particularised knowledge in regard to the entities he handles. He has to measure phenomena of higher orders. He has to particularise his models. In an equation of price and demand such as $d = ap + b$, he wants to know what precisely are the values of a and b. And that also after he is satisfied that the general form of the demand equation is of the form given by $d = ap + b$. In other words, he wants

to know the *structure* of the equation and then the precise nature of its components.

For knowledge of the above type which he needs in order to particularise his models an applied economist has to make use of the science of statistics which, for his purpose, has to be more or less mathematical. After pure theory has suggested some broad features of a general model the applied economist has to go to the market place with his knowledge of statistics for what is called statistical inference. He has to satisfy himself that the general form or structure of the equations he has borrowed from the pure theorist is correct or sufficiently correct for his purpose.

In this process (of statistical inference), the first important step is that of measurement of second-order phenomena. The statistican calls this estimation. He has to estimate the values of symbols. For both, the determination of the general form (the structure) of the equations and the evaluation of the coefficients, he has to measure the entities involved. There are statistical methods that he has to adopt for this purpose. He has to go to the market place or rely on data already gathered from the market. He simplifies his work by making generalisation: he observes a few well-selected cases and then applies his findings to the whole universe. This process is called sampling. We are not statisticians and so we do not need to go into these details. Suffice it to note that statistical methods exist for the purpose of making estimates, i.e., for the purpose of measuring phenomena with the object of particularising a general model.

One can always make mistakes. No one is infallible; we are all imperfect. An applied economist is, therefore, never sure that he has got a correct general model and has been able correctly to particularise it. He has, therefore, to subject his models to the process of verification. He has to detect errors and then to minimise them. He can never get rid of all errors and so he has to do the second best thing, namely, to minimise them. He discovers limits within which the errors must lie.

Types of Structural Equation

A model must consist of at least two equations, but there is no upper limit to the number of equations constituting a model. When the universe of our investigation is wide and complex we have naturally a large number of entities to handle: the number of variables becomes large. And when the variables are large in number we have to have an equally large number of equations. These equations, structural as they are called, differ in respect of the relationships between variables that they indicate. Some show relationships which are due to technological features of an economy; others show relationships that are due to natural factors. Some relationships are due to habits of behaviour of people and

others, again, are due to the social or institutional framework of an economy. Accordingly, equations are caleld technological, natural, behaviouristic and institutional. And, added to these, we also have equations that are called definitional; they equate two things which are made equal by definition. For example, we define income as consisting of expenditure on consumption-goods and expenditure on production-goods, i.e., investment. And such a definition of income gives us the equation $Y = C + I$, where Y is income, C consumption and I investment. Such an equation is called a definitional equation.

All these types of equation define the properties of an economic system. And when its properties are defined, we have a model which we can use for whatever purpose we want to. These equations give us the broad features of an economy. They are broad because, in the first place, they relate to macro-entities and, in the second, because they indicate no precise quantitative relationships between the various entities of the system. To reduce their broadness, or to particularise them, we have to find the numerical values of the algebraic symbols. This is the work of estimation we have referred to above.

Determination of the Values of Coefficients

If we have to draw a curve to represent the relationship between price and demand we must know from which point the curve starts and how it proceeds from left to right. In other words, we must know the position (location) of the demand curve and its shape. In the case of a linear (straight line) demand curve of the form $d = ap + b$, if we know the value of b we know from what point on the axis of y the demand curve starts and if we know the value of a we know its shape (the slope of the line).

The coefficient a in the above equation is the index of the elasticity of the demand curve. We know that elasticity of demand is measured by $-f(x)/x.f'(x)$. And, in the above equation, it is equal to $-p/ax$. Thus the elasticity of demand at any point is dependent on a. The necessity of determining the value of a can be described, in other words, as the necessity of knowing the elasticity of demand.

The coefficients and the parameters of an equation constitute its structure. In the above example of the demand curve b and a are the given constants. They define the structural form of the equation. While they are constant, d and p are variables. The demand d depends on the price p and, conversely it can be said, p depends on d. These variables determine and are determined by each other. This is why they are called endogenous variables. But a and b are (given) constants: they can be treated as exogenous variables. We can call them variables on the ground

that they can also change; nothing is permanently constant in a world that is changing. They are either assumed to be constant or taken to be so for the time being. And most generally, they are constant in the important sense that changes in d and p do not *directly* do anything to cause a change in a and b. For the purpose of the determination of the values of d and p these entities are taken to be constant.

To come back to the previous point, we have to determine the values of the coefficients. In the above demand equation there is only one coefficient, namely, a. If we know the value of the constant b, the only other thing we need to do is to find out the value of a. And since b is also not actually fixed we have to assign a value to it. All this the econometrician has to do. This is the work of estimation for which he has to use statistical methods which he applies to data obtained from the market place.

But in the market place the type of information we need is not found ready-made. All kinds of forces are mixed up and all kinds of changes in all the variables are found intermixed. The one important and difficult task for the econometrician is to separate these various influences one from another. Take, for instance, the relation between price and demand. In the demand equation that we have used above demand is shown as a function of price, meaning that it changes as the price changes. But there is also the constant element b in the equation. It determines the location of the demand curve. In the market this influence b is not constant for the simple reason that nothing is constant there. Our data are gathered from events that have occurred over a period of time. And during that period all influences have changed. Forces that change a, forces that change b and forces that change the price all act concurrently. From our point of view, some of these variations are endogenous and some exogenous. It is easy to find the value of a if and when b does not change. But since it does change the task of the econometrician becomes difficult though not impossible.

Methods Employed by some Econometricians

Let us take demand and supply equations of the simple linear types with random variables to take account of uncertainty. Such examples have been taken and the solution indicated by Professor Wassily Leontief and others. Let us put down the demand and supply equations as follows:

$$D(t) = a'p(t) + \beta y(t) + u'(t)$$
$$S(t) = a''p(t) + u''(t)$$

Here demand is shown as a function of price (the type of relationship being given by a') and income. The manner in which income influences

demand is indicated by β. Thus there are two coefficients a' and β. Further, demand is dependent on an uncertain influence called the random variable, u'. We have said earlier something about random variables and so we shall not repeat it here. Supply is shown as a function of price and a'' indicates the type of influence that price has on supply. The random element in the supply equation is u''. Supply does not directly depend on income and so there are no terms with y in the supply equation. To express this in words which would please a mathematician, the coefficient of y in the supply equation is zero.

The slopes of the supply and demand curves (roughly, their elasticities) are determined by the values of a' and a'', while their location is determined by the random elements u' and u'' together with βy. In equilibrium, demand has to be equal to supply. We have, therefore, to put $D(t)$ equal to $S(t)$ in order to determine the price, supply and demand. Let us repeat, it is only equilibrium values of variables that can be determined. This point is so important that it bears frequent repetition. We can *determine* the position of equilibrium and no other position. Determination consists in finding the norm, the norm of any magnitude. So, equating demand and supply and subjecting the equation to the process of simplification we get values of price and demand (which is equal to supply). They are:

$$p(t) = \frac{\beta}{a'' - a'}\,y(t) + \frac{u'(t) - u''(t)}{a'' - a'}$$

$$D(t) = \frac{a''\beta}{a'' - a'}\,y(t) + \frac{u'(t)a'' - u''(t)a'}{a'' - a'}$$

The pure economist is satisfied with such values of price and demand. He does not need to know what a', a'', β, u' and u'' are. He lives on a theoretical plane; but an applied economist lives in the market place, as it were. He must know what these algebraic symbols are equal to. That is why he needs to use statistics and subject his data to statistical analysis. He takes the statistics of price and income and plots them with reference to two axes. He then draws a curve through these points. The curve he selects is the one that has the closest fit. He uses the method of *least squares* which ensures that the curve traced out harmonises as much with the figures of price and income as possible.

Enough has been said to indicate the lines on which econometricians proceed to get detailed information about the material they handle. Let us repeat that the pure theorist needs to measure first-order phenomena while the applied economist has to be able to measure second-order and higher-order phenomena. While a pure theorist is contented with *economic* laws of supply and demand the econometrician is not satisfied till he has been able to derive *statistical* laws.

Application of Econometric Methods

The econometric methods outlined above have been used in particular for the derivation of demand function, supply function, cost function and production function. As we observed above, to be able to derive a function we must know the shape and position of a curve. The position is given by structural parameters and the shape by the coefficients of variable entities. Speaking in more particular economic terms we can say that for the determination of the shape of a demand, supply, cost or production curve we must know its elasticity. Econometric methods are used to determine the structural parameters and the elasticities of curves.

Further, econometric methods have been used to construct macro-economic models of an economy. Such models can be static or dynamic. The models built by pure economists are all in terms of algebraic symbols which are of little use for an applied economist. Econometric methods are used to find, by the application of statistical analysis, numerical values for those algebraic symbols.

Econometric methods have also been employed to find utility functions of indifference systems. In all such processes the statistician has to rely on observation or observed facts. He can, to an extent, depend on his ability to experiment. But the scope for experiment is strictly limited; for, an economic statistician has little control over the environments within which his operating units function.

PART THREE

Modern Features of Economics

CHAPTER 10

Theory of Games

Introduction

THE 'theory of games', 'game theory' and 'theory of games of strategy' are used as synonyms for a study which looks upon human behaviour as characterised by conflicts. These conflicts are classified dichotomously as being due to (*i*) irrationality, or (*ii*) rational, conscious artful behaviour. The latter type of conflict is called a *contest* or a *game of strategy*. The word *strategy* is intended to focus on the interdependence of (two) parties in contest on their *expectations* of each other's behaviour. The contestants are called *players*. The best course of action or *move* for a player depends on what the other player is expected to do. Each player of a game wants to *win*, if necessary by outwitting other players.

Games of strategy are distinguished from games of skill and games of chance. In the ultimate analysis, they are not different: chance, which characterises games of chance, comes in through expectations in games of strategy; and skill, in games of skill, implies wise decisions in the face of different expectations.

Games of strategy assume rational behaviour to reach the *best* result. 'Best' does not imply *only* 'best' for a player: it does not exclude the motive of the complete extermination of other players: and it does also have a place for a 'win' through bargaining, mutual accommodation and avoidance of mutually damaging behaviour. In fact, usually in a game, as in actual life, there is implicit a powerful common interest.

Even in a worker's strike success does not mean a success that destroys the employer.

When economics is defined as a study of optimal allocation of scarce resources, optimality is taken to imply *maximisation* of some objective by an all-knowing, omniscient, lightning-quick calculating, rational man. But it should not be forgotten that the individual decision-maker has limited time at his disposal, limited perception, limited knowledge, including knowledge of only a few alternatives, limited computing ability: and his alternatives carry different (not exactly known) degrees of risk. He may even not know exactly what he wishes to maximise and his ability to maximise may be hampered, if there are more players to a game, with goals that are not identical. So, while small decisions may be made rationally, decisions regarding huge projects and monuments may be made by the heroic man without calculation of cost.

In this background, games of strategy are being developed partly to help the practical (business) man and partly to throw light on dark corners in the economic theory of bilateral monopoly, duopoly and oligopoly. When one cannot implement his maximisation ideal one takes to a policy of deterrence: he avoids certain (undesirable) consequences or disasters by avoiding the related courses of action in the face of a *threat* of external force. Such a view finds use even in military affairs, criminal law and child-discipline and may be said to lie at the root of a wealthy paternalistic nation's threat to stop aid to a poor nation with a weak and disorganised government unless it follows sound economic policies. To give a simple illustration, let there be 3 possible circumstances and 4 possible actions and let the following table indicate the consequences of each combination of circumstance and action:

Actions	*Circumstances*			*Minimum*
	C_1	C_2	C_3	*results*
A_1	200	78	0 (= Death)	0
A_2	112	1 (=Disaster)	171	1
A_3	98	109	99	98
A_4	171	164	0 (= Death)	0

Suppose we know that circumstance C_2 is certain: then rational decision shall be in favour of action A_4. If this knowledge is not available but we know that C_1, C_2 and C_3 occur with probability p_1, p_2 and p_3, then we may use the mathematics of probability and calculate the expected value for each action. Thus, the expected value for A_3 will be given by

$$E_3 = p_1(98) + p_2(109) + p_3(99)$$

So, knowing E_1, E_2, E_3 and E_4, rational behaviour may be to favour action to ensure the highest of these four expected values.

Suppose probabilities are also not known: but we know that C_3 implies 'death' under actions A_1 and A_1, and C_2 implies disaster under A_2: only under action A_3 are there no deaths and disasters. What will rational behaviour now prompt one to do? Will the decision-maker try to avoid the disasters and deaths and hence decide against A_1, A_2 and A_4 even though these would yield higher results if C_1 and C_2 happen to materialise? In more general terms, a policy of deterrence may be followed and the decision may be to consider the least or minimum results (0, 1, 98 and 0) of the actions and decide upon action A_3 which ensures the maximum of these minimum results. Such a decision is called a *maximin* decision.

This does not mean that a policy of maximin decision is followed by all players under conditions of uncertainty. There are individuals in this world who are more afraid of having to regret later for not taking the chance of bagging a good thing—in our example to get 200. There are also individuals who would act for a heavy prize even though many, who tried for it in the past, might have failed. No doubt, in every case the action may be said to be in accord with the best satisfaction of the decision-maker: only the criterion of 'best' changes. We might say that one who *maximins* his gain is also following the maximisation principle in an extended form.

Essentials of a Game

Let us now understand a game with reference to duopoly.

Let there be two producers, A and B. Each has a number of moves to choose from. Say, A can select (*i*) a particular price, or (*ii*) a particular price and a particular advertisement expense, or (*iii*) a price, an advertisement expense and a form of advertisement (e.g., newspaper, radio or handbills). Similarly, B has to make a selection. The selections are called the 'strategy of A' and the 'strategy of B'. For each combination of the strategy (of A and B) the consequence is known: it is known how much A will gain and how much B will gain. The gains of A (or B) are called A's (or B's) *pay-offs*. The pay-offs will no doubt depend on market and demand conditions.

This game is called a *two-person game*. In general, there can be 1, 2, 3, . . ., or n persons games.

If the total of pay-offs of A and B for each strategy-combination is the same, the game is called a *zero-sum game*; otherwise, it is called a *non-zero-sum game*.

Let us consider a numerical 2-person zero-sum game for our duopoly problem with the following *pay-off matrix*:

A's *strategy*	A's *pay-offs* B's *strategy* P_1	P_2	P_3	B's *pay-offs* B's *strategy* P_1	P_2	P_3	*Minimum* A's *pay-off*
Raise price (P_1)	27	18	25	73	82	75	18
Price unchanged (P_2)	50	5	95	50	95	5	5
Lower price (P_3)	30	64	12	70	36	88	12
Maximum A's pay-off	50	64	95				

Usually, in such cases, *only* one player's pay-off matrix is given: conventionally, it is of the player whose strategies are given in the first column.

The last column shows the minimum pay-offs or gains of *A* for each of his strategies: these are the minimum value in each row of *A*'s pay-offs.

Similarly, the last row shows the maximum gains of *A* for each strategy of *B*.

To repeat, we have here—

(*i*) a fixed number of players;
(*ii*) a fixed finite number of strategies for *each* player;
(*iii*) a known consequence of each strategy-combination of players;
(*iv*) a knowledge on the part of *each* player of the strategies and consequences;
(*v*) a will on the part of *A* to maximise his gain and in any case to avert reduction in gains due to the other's move;
(*vi*) a will on the part of *B* to minimise *A*'s gain and in any case to avert increase in *A*'s gain;
(*vii*) a state of knowledge on the part of *A* (as also *B*) of the possible strategies of the other player;
(*viii*) a will on the part of *A* (and *B*) to find (by expectation) the strategies of the other player and to outwit him.

Now the greatest (i.e., maximum) value out of all minimum possible gains (18, 5 and 12) of *A* is 18 and so *A*'s move will be P_1: it is his *maximin strategy*.

On the other hand, the lowest value out of all maximum possible gains (50, 64 and 95) of *A* is 50, and so *B*'s move will be P_1: it is his *minimax strategy*.

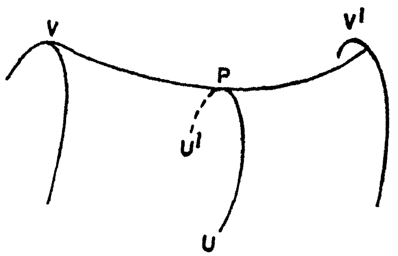


Diagram 10.1

As a result, the solution on first moves will be a gain of 27 which is at the cross-road of P_1 and P_1 in relation to *A* and *B* respectively.

If a player's maximin-strategy-consequence is the same as the other player's minimax-strategy consequence, then the consequence is called an *equilibrium point* or *saddle point*. It

possesses an element of stability in the sense that both players are satisfied at having followed their strategy successfully. The name 'saddle-point' is on the analogy of a horse's saddle where the point P is both the maximum point on curve UU' and the minimum point on curve VV' (Diagram 10.1).

In the above example 27 is not a saddle point. A can argue to himself that, knowing this strategy of B, he could have improved his gain by adopting strategy P_2 or P_3. As it is, A will be said to be *pleasantly surprised* at getting 27 instead of 18.

We must note a number of points:

1. The policy of choosing maximin strategy by A ensures the largest share of the gain (or *market*) which B cannot reduce further. Similarly, B's minimax strategy enables B to prevent A from increasing his share further.
2. The maximin-minimax point for A is also the maximin-minimax point for B on the basis of B's pay-off matrix.
3. A maximin strategy is a poor counter-move to a non-minimax strategy. Thus, in our example, if A follows maximin strategy and B does not necessarily follow a minimax strategy, A stands a chance of getting 18 or 25.

Not only the duopoly problem but also the oligopoly problem may be seen as a two-person game by grouping all competitors in one body. But it is not easy to assert that the game of strategy is an apt model of oligopoly or even duopoly.

Mixed Strategy

There is, incidentally, an aspect of the game of strategy which may be mentioned. If A knows the strategies of B and if he decides to select his strategy 'statistically', i.e., with given probabilities, he can sometimes determine these probabilities such that the expected value under each strategy of B would be the same. Thus, if the probabilities be p_1, p_2 and p_3, then, for the first strategy of B, the expected value (λ) (by definition) will be given by

$$p_1(27) + p_2(50) + p_3(30) = \lambda$$

Similarly, there will be two more equations:

$$p_1(18) + p_2(5) + p_3(64) = \lambda$$
$$p_1(25) + p_2(95) + p_3(12) = \lambda$$

These can be solved in terms of say $\dfrac{p_1}{p_3}$ and $\dfrac{p_2}{p_3}$;

$$\frac{p_1}{p_3}(27) + \frac{p_2}{p_3}(50) + 30 = \frac{\lambda}{p_3}$$

$$\frac{p_1}{p_3}(18) + \frac{p_2}{p_3}(5) + 64 = \frac{\lambda}{p_3}$$

$$\frac{p_1}{p_3}(25) + \frac{p_2}{p_3}(95) + 12 = \frac{\lambda}{p_3}$$

Putting $\frac{p_1}{p_3} = \lambda_1$ and $\frac{p_2}{p_3} = \lambda_2$ we can write—

$$27\lambda_1 + 50\lambda_2 + 30 = 18\lambda_1 + 5\lambda_2 + 64$$

$$25\lambda_1 + 95\lambda_2 + 12 = 18\lambda_1 + 5\lambda_2 + 64$$

or,
$$9\,\lambda_1 + 45\,\lambda_2 = 34$$

$$7\,\lambda_1 + 90\,\lambda_2 = 52$$

$$\therefore \qquad \lambda_1 = \frac{16}{11} \text{ or } \frac{720}{495};\ \lambda_2 = \frac{230}{495}$$

Hence the probabilities are $\frac{720}{1445}$, $\frac{230}{1445}$ and $\frac{495}{1445}$ and these will yield the same expected value whatever be the strategy of *B*.

For this policy also, we must know the *pay-off matrix* and it must remain unchanged. Hence, the market demand and the demand price must be given and fixed, and the profit must also be the same per unit of sale: otherwise, the pay-off matrix will change as a consequence of changing demand conditions and cost-conditions due to different organisation and technique. Nevertheless the choice of a strategy statistically may well appeal to a firm or business which wants to keep the competitors (i.e., other players in the game) guessing.

Also, by properly varying the probabilities, sometimes the maximin expected values can be improved upon those given in the pay-off matrix. And so, we can easily conceive of an *optimal mixed strategy*, i.e., a best mixed strategy which maximises the expected values. And if both *A* and *B* determine their optimal mixed strategies it can be mathematically shown that in a zero-sum game both will mean the same equilibrium gain (λ) to *A*.

The original duopoly case is said to be a game of *pure strategy* and the one involving *statistical choice of strategy* (with probabilities) is said to be a game of *mixed strategy*.

Game Theory and Linear Programming

Incidentally, the problem of optimal mixed strategy may be looked upon as a linear programming (see the next chapter) problem. In our example, we may say that we want to maximise λ subject to the conditions—

$$p_2 \geqslant 0;\ p_2 \geqslant 0;\ p_3 \geqslant 0$$
$$p_1 + p_2 + p_3 = 1$$
$$27p_1 + 50p_2 + 30p_3 - \lambda \geqslant 0$$
$$18p_1 + 5p_2 + 64p_3 - \lambda \geqslant 0$$
$$25p_1 + 95p_2 + 12p_3 - \lambda \geqslant 0$$

Games in Extensive Form

In this world there is seldom a given pay-off matrix, much less one for all time. Also, decisions are taken (i.e., moves are made or strategy selected) again and again. Every past experience may affect the next move. Thus, as soon as A finds that B's strategy was P_1 he may expect that B will repeat it the next time as well and he may change his move from P_1 to P_2. Similarly, B may grow wiser as a result of A's last move and may choose strategy P_2 to reduce the pay-off from 27 to 18. As a result, A's pay-off will be reduced from 27 to 5. We may continue this study of a game, move by move. This is called a study of a *game in extensive form*, and our original case, a study of a *game of normal form*.

Since a game in extensive form, though nearer actual conditions, means changing fortunes for both A and B, they may not like to run the risk and hence may like to *play it safe* and take a decision once for all on the basis of the pay-off matrix.

Non-zero-sum Game under Collusion (or Coalition)

We said in the beginning that even in a worker's strike there is a powerful common interest: a successful strike does not mean complete extermination of the employer. This idea may easily lead one to think of *collusion* (or *coalition*) between the two players in a two-person game or between two or more-players in a three or more-person (i.e., n-person, where $n \geqslant 3$) game. By collusion two players may well increase their total profit.

To give a simple example from mathematical economics, we know that, for a given demand-price relation

$$p = -\alpha x + \beta$$

and given total cost function

$$T = Ax^2 + Bx + C$$

the equilibrium output, assuming that each duopolist maximises his profit taking the other's output to be constant, is given by

$$x_1 = x_2 = \frac{\beta - B}{3\alpha + 2A}$$

and each producer's profit will be

$$(\alpha + A)\, x_1^2 - C \text{ or } \frac{(\alpha + A)}{(3\alpha + 2A)^2} \cdot (\beta - B)^2 - C$$

On the other hand, if both combine, it becomes a case of a monopolist with two identical factories. Then the equilibrium output is given by

$$x_1 = x_2 = \frac{\beta - B}{2\,(2\alpha + A)}$$

and each producer's profit is given by

$$(2\alpha + A)\, x_1^2 - C \text{ or } \frac{(\beta - B)^2}{4\,(2\alpha + A)} - C$$

This second profit is greater as

$$\frac{1}{4\,(2\alpha + A)} > \frac{\alpha + A}{(3\alpha + 2A)^2}$$

So, if the players (or producers) can make full cooperative use of all opportunities yielding mutual advantages, they will do so. According to Baumol "they will always end up somewhere on the *contract curve*." In doing so, there will lurk a danger: they may not agree on how to distribute the total gain (or, spoils!) resulting from collusion. There can be put forward criteria of a fair division but in a dynamic world influenced by selfish motive, unless one firm merges with the other, the collusion may break on the rock of the criterion for division of gain.

Non-zero-sum Game under Non-collusion (or Non-coalition)

If the players do not take recourse to collusion, they may choose strategies which will decrease the gains of both. Or, although they do not co-operate, either may advertise his next move (say, to increase price) so that others may also act likewise and the market may be subjected to joint exploitation. This will be a form of tacit collusion. Alternatively, a player may force the other to yield by publicising a threatening action in case he does not win; and if his threat goes home, he may improve his pay-off matrix, and his equilibrium pay-off. Yet, if out of sheer fear of loss, A acts in a (maximin or any) way while B does not, A's move together with B's move (out of similar considerations) may mean disaster to both and a reduction in total economic welfare.

n-Person Game

The theory of games is of little use at this point: really, as we try to

approach actual conditions, the theory of games rapidly recedes in efficiency. Actual conditions demand not only the non-zero-sum condition but also an n-person's situation. If the number of players becomes more than two, *collusion* (*or, coalition*) becomes more feasible and likely. There may even be '*side-payments*' to certain players to keep them out (or, inactive). In any case, the share of each party to collusion must not be less than what he could get himself by acting independently; but, for the success of the collusion, the pay-off shares must be such as would ensure the maximum *coalition-gain*. This is difficult to achieve. Besides, it is difficult to appraise the change in bargaining powers when a player succeeds in securing a coalition with another player. Actually, at one time (1961) there were twenty-odd theories of behaviour in n-person games and none of them was universally satisfactory. Suffice it to say that in this regard the development of the theory is little fruit-giving: it is, to some extent, light-giving.

Limitations in General

Finally, the theory of games does not avoid the maximisation urge: it only emphasizes consideration of the maximin motive also. For the theory of games, a knowledge of the strategies and the related pay-off matrix must be available and known beforehand, which has very little feasibility in practice. The assumption that each player knows the strategies of the others is also of doubtful feasibility. Lastly, when it is said that a player will tend to outwit the others, he should logically get involved in an endless chain of thought to which he is not likely to yield in practice.

This last point deserves to be explained a little (say) with reference to our A-B pay-off matrix. A may argue as follows by stages:

(*i*) Let me choose P_2 because B will choose P_1.

(*ii*) B will be able to see from the pay-off matrix (also known to B) that I will choose P_2 and so he may choose (say) P_2: in view of this I should choose (say) P_3.

(*iii*) B will be able to see this too and will therefore choose strategy P_3: so I should choose P_2.

(*iv*) But B can also think this much out and he is then likely to choose. . .There is no end to the chain if we admit that each firm (or player) tries to outwit the competitors (or other players).

Contributions of the Game Theory

Three contributions deserve mention:

(*i*) The game theory provides us with a new angle of study of certain mixed economic problems, particularly those of oligopoly and duopoly, and emphasizes the importance (normative aspects) and the likelihood of the urge (positive aspect) to seek a *maximin* result.

(*ii*) This theory emphasizes the importance in economic life of *coalitions* and *side-payments*.

(*iii*) We get the concepts of statistical choice of strategy, usually called *mixed strategy*, and of *optimal mixed strategy*.

(*iv*) The theory is useful for teaching and training purposes for developing appreciation of "inter-relationships between different functions in an organisation and for giving individuals deeper insight into complex problems of interaction."

CHAPTER 11

Linear Programming

Introduction: Setting the Stage for Linear Programming

Let us consider for a moment how analysis is usually done in economics. A producer produces a commodity. We assume the commodity to be divisible and, all units of it, homogeneous. He produces it by using factors of production which, in economics, are divided into four (or five) categories. The relationship between output and factors of production is indicated by a production function. Now we emphasize (1) that the number of producers producing more than one commodity is increasing; (2) that a producer chooses one (or some) out of a finite number of production processes or methods of production; (3) that for each production process there is a fixed pattern of use of factors which may be easily classified into more categories (called inputs) than the limited 4 (or 5) categories of the old, because now mathematics enables us to consider them for applied problems; (4) that a producer does not necessarily use the whole of all inputs available; (5) that the producer's decision is based on money-valuations generally; (6) that the expression giving possible money-valuation does not yield a continuous curve (or function) due to the indivisibility of the commodity produced so that the first and second order partial derivatives are not feasible. Hence we see that the usual marginal analysis is not applicable.

Besides, marginal analysis assumes that a position of maximum net revenue (or, optimum) is a position of equilibrium. A firm, it is submitted,

attempts to maximise its profits (as measured by some quantity) over its (relevant) whole life. The measure of this profit is *taken* by *some* to be the present value of the expected series of profits discounted at market rate of interest or some internal rate of return. Others question this and submit that, like a consumer, a firm also takes a decision in consideration of (not merely profits but also) prestige, good relations (with employees, customers or the public), security, liquidity and peaceful working. Still others have been of the view that a firm estimates an average cost and adds a percentage (called *mark-up*) of cost for profits, presuming that all output will be sold out. Finally, it is questionable whether (as envisaged under marginal analysis) a firm has all the information and considers all this information in taking a decision. Given perfect competition, it can be assumed with some justification that average cost is constant and known and that price is known. But when one passes to consider non-homogeneous production, non-constant costs and imperfect markets, the information problem becomes difficult, particularly because we want to know now not only what is but also what may be. Uncertainties about production (as in agriculture) and future prices exist to an extent under perfect competition also but these increase considerably under imperfect competition.

Varieties of Programming

Decision-making under imperfect knowledge has in recent decades led to the admission to economic theory of games theory, input-output analysis, linear programming, and organisation theory. We still conceive (*i*) of a method of selecting the best position among possible positions by maximising (or minimising) an expression, and (*ii*) of a limited field of choice of possible positions. Not all these new theories were primarily developed for economics. Linear programming was particularly developed as a technique of planning the diversified activities of an air-force.

We now have mathematical programming, linear programming, integer programming, quadratic programming and non-linear programming. Mathematical programming is a general label used in the business world more or less as a synonymn for 'linear' and 'integer' programming, but it should preferably be used to cover all the subsequent four labels. The other four labels are not all mutually exclusive. Quadratic programming may be considered to be a part of non-linear programming, which in turn is mutually exclusive to linear programming. Linear programming is a technique of study wherein we consider the maximisation (or minimisation) of a linear expression (called the *objective function*) subject to a number of linear equalities and inequalities

(called *linear constraints*). The final solutions need not be all in the form of integral values. Where we want integral values only, we slip to the field of integer programming. Where our objective function and other relations are non-linear our study is covered by non-linear programming.

A Simple Illustration

Given an objective and a number of resources, finite in number and quality, the objective may be achieved by using some of these resources in full and others in part. The resources used can *never* exceed what are given. Thus the amount of any resource used is either equal to or less than the amount available. Suppose an objective can be achieved through the use of factor X and factor Y in different quantities (x and y) as indicated below:

$$\begin{aligned} (i) \quad x &= 7, \quad y = 12 \\ (ii) \quad x &= 10, \quad y = 4 \\ (iii) \quad x &= 20, \quad y = 1 \end{aligned}$$

Further, suppose that $2x + y$ cannot exceed 25 due to finiteness of resources. Then, of the three choices, the first and the third must be rejected as impossible (or, not-feasible) since—

$$\begin{aligned} (i) \quad &\text{If } x = 7, y = 12;\ 2x + y = 26 \\ (ii) \quad &\text{If } x = 20; y = 1;\ 2x + y = 41 \end{aligned}$$

Only the second choice is feasible since, when $x = 10, y = 4$,

$$2x + y = 24.$$

There may be a number of such feasible solutions and we may choose one that minimises or maximises another linear function of cost or profit.

We can illustrate the above example geometrically (Diagram 11.1). We may draw the straight line

$$2x + y = 25$$

Any point in the positive quadrant XOY represents a combination of factor X and factor Y. From geometry we know that if we draw the straight line

$$2x + y = 25 \text{ } or \text{ } 2x + y - 25 = 0$$

then any point P with coordinates (h, k) which lies above it has co-ordinates such that

$$2h + k - 25 > 0$$

$$\text{i.e.,} \qquad 2h + k > 25$$

Hence, all combinations above the AB line are not feasible. Our objective therefore has to be achieved by choosing a point, i.e., a combination of factor X and factor Y, lying in the shaded area, including the boundary

lines *OA*, *OB* and *AB*. Lines *OA* and *OB* can be justifiably included if we bear in mind and introduce the assumption that the quantity used of any of the factors *cannot* be negative, i.e.,

$$x \nless 0; \quad y \nless 0$$

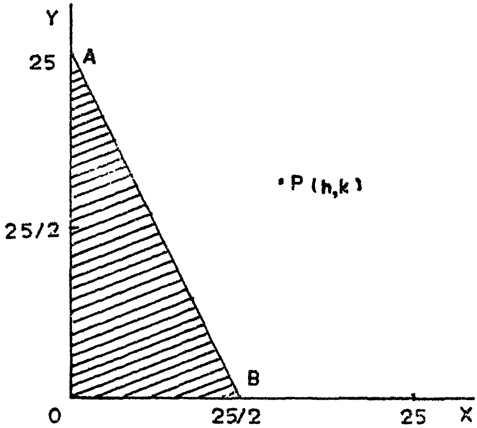


DIAGRAM 11.1

These are called *conditions of non-negativity* of x and y.

So $x = 0$ and $y = 0$ become border lines.

The shaded area constitutes all feasible solutions and a point has to be selected in it as a linear programming solution with a view to minimising or maximising some linear function.

Let us understand the use of such an approach by a simple example each from consumption, production, and distribution.

A CONSUMPTION PROBLEM

Let C_1 and C_2 be two items of consumption, each containing two nutrients. Let a kilogram of C_1 contain 2,000 calories and 40 grams of protein; and a kg. of C_2 contain 4,000 calories and 160 grams of protein. Let a balanced diet contain 2600 calories and 80 gms of protein. Let the market price be 60 paise per kg. of C_1 and Rs. 1·50 per kg. of C_2. What quantities of C_1 and C_2 should be selected to minimise the cost and attain the balance of diet?

Let us choose 1,000 calories and 20 gm. as units of measurement for calories and protein, a kilogram as the unit of weight and 30 paise as the unit of price. Then we may tabulate the data as below:

	One unit of commodity C_1	C_2	Requirements of balanced diet
Contents:			
Calories	2	4	2·6
Protein	2	8	4
Price	2	5	

If finally the consumer chooses x units of C_1 and y units of C_2, then the cost (Z) will be given by

$$Z = 2x + 5y$$

We have to minimise this objective function subject to the condition that x and y are not negative and that the calorie and protein intake is at least the minimum required. Therefore

$$\begin{aligned} x &\nless 0 \\ y &\nless 0 \\ 2x + 4y &\nless 2{\cdot}6 \\ 2x + 8y &\nless 4 \end{aligned}$$

Here the objective function (Z) to be minimised is a linear function of x and y; and the inequality conditions also involve only the first (linear) powers of x and y. Hence it is said to be a problem in linear programming.

If we were to go in for achieving a balanced diet exactly, we could write

$$2x + 4y = 2{\cdot}6$$
$$2x + 8y = 4$$
$$Z = 2x + 5y$$

The problem of finding x and y in order to minimise Z will not arise here. The first two relationships (or, equations) are sufficient to determine x and y.

The first two equations are said to be *side-relations* and one of them is sufficient to study how to minimise Z. We may use (say) the first equation and write

$$\begin{aligned} 2x &= 2{\cdot}6 - 4y \\ \therefore \quad Z &= 2x + 5y \\ &= 2{\cdot}6 - 4y + 5y \\ &= 2{\cdot}6 + y \end{aligned}$$

and then we could say that Z is minimum when $y = 0$. Hence $x = 1{\cdot}3$. But this will not ensure 4 units of protein, it will give only $3(1{\cdot}3)$ or $3{\cdot}9$ units of protein.

A similar unbalance may be found if we were to use only the second side-relation.

But there is a way to solve the problem subject to both side-relations. Diagram 11.2 gives a graphic exposition, which is easier to understand.

We plot the relationship

$$Z = 2x + 5y$$

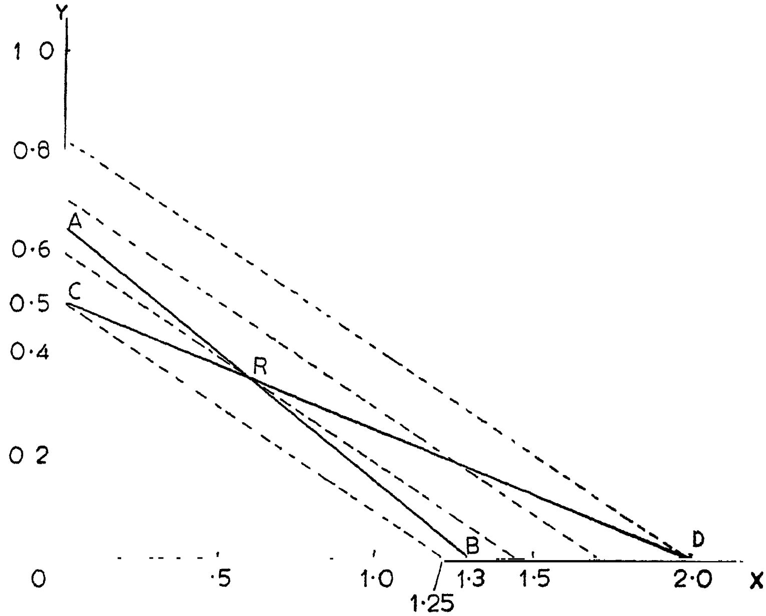


DIAGRAM 11.2

for different values of Z. We get a family of dotted lines parallel to each other. Each dotted line corresponds to a finite value of Z. We must choose the lowest of these such that the requirements of a balanced diet are at least satisfied.

Now, if we plot the relationship

$$2x + 4y = 2{\cdot}6,$$

we get a straight line *AB* intersecting the axes in *A* and *B* as shown. Then any choice-point *P* lying to the right of *AB* will be such that units of heat (calories) received will not be less than 2·6. So any final choice has to be either on *AB* or to its right.

Similarly, we may consider $2x + 8y \nless 4$ and plot

$$2x + 8y = 4$$

We get the straight line *CD*. Then the choice should be on or to the right of *CD* also.

If *AB* and *CD* intersect at *R*, then, in order to satisfy both side-conditions, the final choice should be on or to the right of the boundary section *ARD*.

So, we must choose that dotted line which is lowest, i.e., nearest to the origin *O* and yet has a point common with the boundary line *ARD*. As it is, this is the dotted line passing through *R*, which point therefore indicates how many (i.e., 0·6) units of commodity C_1 will be purchased and how many (0·35) units of commodity C_2.

It is easy to see that since AR and RD are straight lines, the final dotted line would in general be that passing through either A, R or D. Hence, in general, the final point is either one of the end-points or one of the intersection points along the boundary line. In a particular case the dotted line may coincide with one of boundary sections: then the solution will *not* be *unique*.

All solutions or choices indicated by points on the boundary line ARD and to its right together constitute the *feasible solutions*. The solution at R is the *optimal feasible solution*.

A Production Problem

Let us now take up a production problem.

A firm can produce commodities C_1 and C_2 and the per unit profit is 1·9 and 2 respectively. If the quantities produced are x and y the total profit (Z) will be

$$Z = 1{\cdot}9x + 2y$$

Let each unit of C_1 involve use of 2 man-hours and 0·5 machine-hour. Let each unit of C_2 involve 1 man-hour and 1 machine-hour. Let the firm have 4 man-hours and 3 machine-hours available. Then, for producing x and y units of C_1 and C_2 within given resources we must take a decision which satisfies the following two inequalities:

$$2x + y \ngtr 4$$
$$0{\cdot}5x + y \ngtr 3$$

Subject to these two linear inequalities, we have to maximise a linear expression $3x + 2y$ indicating profits. There are the additional conditions that neither x nor y can be negative:

$$x \nless 0; \quad y \nless 0$$

Let us draw the straight line AB (Diagram 11.3) given by

$$2x + y = 4$$

and the straight line CD given by

$$0{\cdot}5x + y = 3$$

Then the feasible solutions must be on or to the left of AB as well as CD. Hence the feasible solutions will be on or to the left of the boundary line CRB.

OC and OB also constitute the boundary lines because

$$x \nless 0$$
$$y \nless 0$$

i.e., because x and y are not negative quantities.

So, the closed area $OCRB$ is the area of feasible solutions.

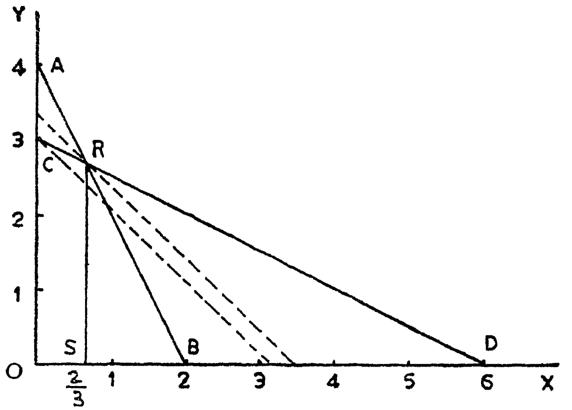


DIAGRAM 11.3

The family of parallel dotted straight lines indicate

$$Z = 1{\cdot}9x + 2y$$

for different values of Z. We have to choose that dotted line which corresponds to the maximum value of Z. Evidently it is the one that passes through the point R. Hence the firm will produce OS of one commodity and RS of the other; this is the optimal feasible solution.

As before, obviously the optimal feasible solution (or, solutions) will be one of the points on the boundary line.

A DISTRIBUTION PROBLEM

In linear programming, a typical transport problem is represented as in the table below indicating sources of supply and destinations. The cells show the cost of transport from each source to each destination. The row-end and column-end figures show the stocks and requirements. (The figures within parentheses and the arrows concern the solution, as explained later.)

Source of supply	*Destination of demand*					*Stocks*
	D_1	D_2	D_3	D_4		
S_1	+ 20 ←	+ 26(4) ↑	+ 14(10)	+ 17(6)		20
S_2	↓ + 10(15) →	+ 15(10)	+ 19	+ 40		25
S_3	+ 20	+ 23(3)	+ 11(12)	+ 31		15
Requirements	15	17	22	6	=	50

We have to choose the most feasible solution to minimise the total cost of transport.

An easy way to get *a feasible solution* is to start allocating stocks to destinations with lowest cost. These are indicated within parentheses in the above table itself.

In keeping with the convention, let cells with initial allocations in parentheses be called *stones* and those without parentheses be called *waters.*

To jump to a *better* feasible solution, a method to serve the purpose of linear programming is to start from a 'water' and then step on stones making a closed path (i.e., reach the original 'water' again) subject to the conditions that

(*i*) no stone is stepped upon twice
(*ii*) no diagonal movement is made
(*iii*) an intermediate water or stone can be skipped.

One such path is shown by arrows starting from the +20 cell.

We then add together all transport costs involved along and round the path starting with that corresponding with the first stone, but change the sign of all even steps. Thus we get

$$+ (+10) - (+15) + (+26) - (+20) = +1$$

It will be realised easily that

(*i*) the number of steps will *always* be *even*;
(*ii*) the cost in the starting 'water' will *always* appear as *positive*;
(*iii*) the magnitude of the total does not change if we reverse the arrows (i.e., the path).

The positive result (+1) of the total is taken to mean that a saving in transport cost will result in this case if we *move* along the path.

The move then is to subtract the smallest positive stone figure from each of allocated figures in cells whose costs were taken without change of sign and add it to the others. Thus we would

(*a*) subtract 4 from 4 and 15, which are stocks for cells whose transport costs were taken without change of sign, and
(*b*) add 4 to 0 and 10 which are stocks allocated to the 20 ('water') and +15 cells.

It will be realised that this means we are shifting 4 units horizontally from the +26 cell to the +20 cell and from the +10 cell to the +15 cell. Hence the change in cost per unit will be

(1) for the first row change: (+20) − (+26)
(2) for the second row change: (+15) − (+10)

which means the total change per unit will be

$$- (+10) + (+15) - (+26) + (+20) = -1 \ ,$$

which appears as +1 in our calculation. Hence the change will mean a reduction of 4(+1) i.e., 4 is transport cost.

Thus the method yields a new arrangement as follows:

	D_1	D_2	D_3	D_4	*Stocks*
S_1	20(4)	26(0)	14(10)	17(6)	20
S_2	10(11)	15(14)	19	40	25
S_3	20	23(3)	11(12)	31	15
Requirements	15	17	22	6	

It will be verified that the total transport cost was formerly $4(26) + 10(14) + 6(17) + 15(10) + 10(15) + 3(23) + 12(11) = T_1$ and it will now be
$4(20) + 10(14) + 6(17) + 11(10) + 14(15) + 3(23) + 12(11) = T_2$
and the saving $T_1 - T_2$ is given on subtraction by
$4(26-20) + (15-11)\ 10 + (10-14)\ 15 = 4(6+10-15) = 4$

The process can be repeated again and again till no closed path leads to further economy.

The method is amenable for use with an electronic computer. And there are modified forms and approximation methods for elementary action. There are methods for taking account of cases where (*i*) resources exceed demand, and (*ii*) demand exceeds resources, so that a part of the demand is bound to be left unsatisfied. But such methods are being replaced by the simplex method.

The Simplex Method

The graphical method or the calculation method used in the case of the distribution problem is not of much help when we come to economic problems where many linear equalities and inequalities arise. For meeting this situation, other algebraic methods have been developed, wherein the necessary operations can be performed with the help of computers, particularly electronic computers.

In order to apply the simplex method the inequalities are converted into equalities by introducing what are called *slack variables.* Thus in our production problem we had

$$2x + y \not> 4$$

and we would introduce a slack (additional) variable, λ_1, equal to the difference between the left- and right-hand expressions so that

$$2x + y + \lambda_1 = 4$$

Similarly, by introducing a slack variable, λ_2, the other inequality would be changed into the equality

$$5x + y + \lambda_2 = 3$$

The expression to be maximised or minimised is always converted into an expression to be maximised. Thus if in our consumption problem

$$2x + 5y$$

is to be minimised, we express it as $O + c_1x + c_2y$ where O is zero, $c = -2$ and $c_2 = -5$ and then maximise it. For when $O + c_1x + c_2y$ is a maximum, its negative counterpart $-(c_1x + c_2y)$ or $2x + 5y$ must be a minimum.

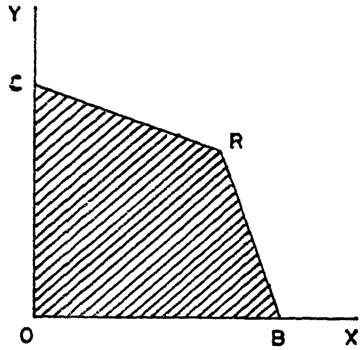


DIAGRAM 11.4

The actual working process of the simplex method is easy to understand with reference to the diagram for the production problem. The feasible solutions were there (Diagrams 11.3 and 11.4) represented by the *feasible region OCRB*, and we had noted that some corner of the feasible region will be indicated as an optimal solution.

As a first step one has to find one of the feasible corners and the *basic solution* corresponding to it. For this solution we calculate the value of the function to be maximised. Say, we do so for the point O; in fact the origin is generally a corner point of the feasible region and is *conventionally* adopted as the starting point.

Next we compute the profits for the adjacent corners C and B. If the profit at (say) B is greater than at O, we move the basic solution to B.

Next we repeat the calculations and do the comparison for B and R: we move to R if the profit is greater for R.

In this way, by *iteration* and successive elimination of points, we arrive at the point of maximum value.

Since with a maximisation expression the feasible region is towards the origin, it invariably contains the origin. Hence a basic solution is invariably given by $x = 0, y = 0$.

The actual numerical calculation can also be illustrated in a simple case. Thus in our production problem we would write

$$\begin{aligned} Z &= 0 + 3x + 2y \\ \lambda_1 &= 4 - 2x - y \\ \lambda_2 &= 3 - 0{\cdot}5x - y \end{aligned}$$

where $x \nless 0$, $y \nless 0$, $\lambda_1 \geqslant 0$ and $\lambda_2 \geqslant 0$.

This is presented in a tabular form as follows:

		x	y
Z	0	3	2
λ_1	4	-2	-1
λ_2	3	$-0{\cdot}5$	-1

Our basic solution is $x = 0, y = 0$ in which case $Z = 0$

Then we interchange (say) the positions of x and λ_1. The idea is slowly to bring x and y into the positions occupied by λ_1 and λ_2. In that case, by making the λ's zero, we can usually get the maximum Z provided the coefficients of λ_1 and λ_2 in the expression for Z

$$Z = \frac{22}{3} - \frac{4}{3}\lambda_1 - \frac{2}{3}\lambda_2 \text{ (see below)}$$

are negative.

However, we rewrite the table after performing operations as below:

$$\because \quad \lambda_1 = 4 - 2x - y$$

$$\therefore \quad +2x = 4 - \lambda_1 - y$$

$$x = \frac{4}{2} - \frac{\lambda_1}{2} - \frac{y}{2}$$

$$= 2 - \frac{\lambda_1}{2} - \frac{y}{2}$$

and

$$Z = 0 + 3x + 2y$$

$$= 0 + 3\left(2 - \frac{\lambda_1}{2} - \frac{y}{2}\right) + 2y$$

$$= 6 - \frac{3\lambda_1}{2} - \frac{3y}{2} + 2y$$

$$= 6 - \frac{3\lambda_1}{2} + \frac{y}{2}$$

and

$$\lambda_2 = 3 - \cdot 5x - y$$

$$= 3 - \cdot 5\left(2 - \frac{\lambda_1}{2} - \frac{y}{2}\right) - y$$

$$= 3 - 1 + \frac{\lambda_1}{4} + \frac{y}{4} - y$$

$$= 2 + \frac{\lambda_1}{4} - \frac{3y}{4}$$

Hence the new table becomes

		λ_1	y
Z	6	$-3/2$	$\frac{1}{2}$
x	2	$-\frac{1}{2}$	$-\frac{1}{2}$
λ_2	2	$\frac{1}{4}$	$-\frac{3}{4}$

from which it can be seen that $Z = 6$ when $\lambda_1 = 0, y = 0$ and, hence, when $x = 2$. This is a better position, a better corner or better feasible solution, than the basic solution $x = 0,\ y = 0$.

Let us now interchange the positions of y and λ_2.

$$\because \quad \lambda_2 = 2 + \tfrac{1}{4}\lambda_1 - \tfrac{3}{4}y$$

$$\therefore \quad \tfrac{3}{4}y = 2 + \tfrac{1}{4}\lambda_1 - \lambda_2$$

$$y = \frac{8}{3} + \frac{1}{3}\lambda_1 - \frac{4}{3}\lambda_2$$

and
$$\begin{aligned} z &= 6 - \frac{3}{2}\lambda_1 + \tfrac{1}{2}y \\ &= 6 - \frac{3}{2}\lambda_1 + \tfrac{1}{2}\left(\frac{8}{3} + \frac{\lambda_1}{3} - \frac{4}{3}\lambda_2\right) \\ &= 6 + \frac{4}{3} - \frac{3}{2}\lambda_1 + \frac{\lambda_1}{6} - \frac{2}{3}\lambda_2 \\ &= \frac{22}{3} - \frac{4}{3}\lambda_1 - \frac{2}{3}\lambda_2 \end{aligned}$$

and
$$\begin{aligned} x &= 2 - \tfrac{1}{2}\lambda_1 - \tfrac{1}{2}y \\ &= 2 - \tfrac{1}{2}\lambda_1 - \tfrac{1}{2}\left(\frac{8}{3} + \frac{\lambda_1}{3} - \frac{4\lambda_2}{3}\right) \\ &= 2 - \tfrac{1}{2}\lambda_1 - \frac{4}{3} - \frac{\lambda_1}{6} + \frac{2}{3}\lambda_2 \\ &= \tfrac{2}{3} - \tfrac{2}{3}\lambda_1 + \tfrac{2}{3}\lambda_2 \end{aligned}$$

The table becomes

		λ_1	λ_2
z	$\frac{22}{3}$	$-\frac{4}{3}$	$-\frac{2}{3}$
x	$\frac{2}{3}$	$-\frac{2}{3}$	$\frac{2}{3}$
y	$\frac{8}{3}$	$\frac{1}{3}$	$-\frac{4}{3}$

from which we get a new solution $x = \frac{2}{3}$, $y = \frac{8}{3}$ for which $z = \frac{22}{3}$ which is greater than the last solution for which $z = 6$.

This is the optimum feasible solution as it has made λ_1 and λ_2 zero and x and y non-negative.

Though we have given this simple illustration, in actual practice electronic computers are used to get the optimal feasible solution.

Integer Programming

Linear programming results need not always be in terms of whole numbers or *integers* while in economic terms goods may be in indivisible units, e.g., locomotives, machines, aeroplanes and trucks. We therefore seek a solution in terms of integers. Where transport problems are involved and places to be visited are indicated, not by distance but by order of visit, a decimal solution such as "visit 9·4th city" has little

meaning. Hence there has been developed what is termed *integer programming.* This technique consists, in essence, in (before seeking the optimal feasible solution) replacing the feasible region *OARB* by a set of points (black dots in Diagram 11.5) which are called *lattice points.* One of these will now be sought as the final solution.

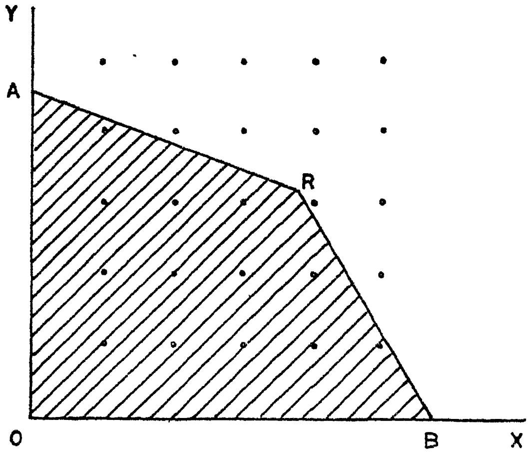


DIAGRAM 11.5

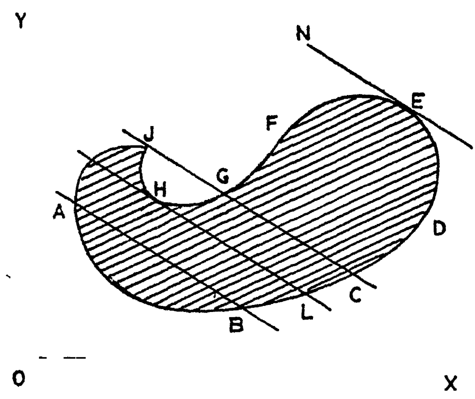


DIAGRAM 11.6

Integer programming is more helpful even in the case of what is called a *non-convex feasible region* (as against a *convex region*). A feasible region (e.g., *OARB*) is said to be convex if a line joining any two points within it lies within the region (*OARB*). A feasible region (e.g., *ABCDEEFGHJ*) is said to be non-convex if when we join (say) points *J* and *G*, the joining straight line is (partly or wholly) outside the feasible region. (Diagram 11.6)

It will be recalled that, in practice, we jump from one point (or, corner) O to another B if, by doing so, the profit increases. And we are sure that it will bring us nearer the optimal feasible solution. But with a non-convex feasible region, if we jump from G to H, and thence to J, we will come to an optimal solution at J which is not the real optimal solution. Although the profit-line is tangential at J, we know from the diagram that a still higher profit line is also tangential at E. So the practical computation may mislead us to a *wrong* optimum feasible solution. In such a case the programming practice does not help us necessarily to reach the final optimal solution.

Certain Working Rules for Linear Programming

In order to render calculations more mechanical, it is necessary to understand certain rules. Given the matrix

$$
\begin{array}{cccc}
 & & x & y \\
Z & 0 & 1{\cdot}5 & 1 \\
 & 4000 & -1 & -2 \\
 & 6000 & -3 & -2
\end{array}
$$

for the first exchange, technically called *pivoting*, choose that column (other than the first column) which has the *largest positive top element*, i.e., (in this example) the x-column.

Next consider rows for which the element in the chosen column is negative. Consider, for each row, the *arithmetic ratio* of the first element to the corresponding element in the chosen-column, i.e.,

$$
\begin{aligned}
4000 \div 1 &= 4000 \\
6000 \div 3 &= 2000
\end{aligned}
$$

and choose the row for which the ratio is smallest; in our example, the last row. The element (-3) at the crossing of the chosen column and chosen row is called the *pivot element.*

Divide (*i*) each of the *other* elements of the chosen column by the pivot element, and (*ii*) each of the *other* elements of the chosen row by the 'pivot-element with its sign changed': change the pivot-element into its reciprocal. This gives us the following new values:

$$
\begin{array}{ccc}
- & -0{\cdot}5 & - \\
- & \frac{1}{3} & - \\
2000 & -\frac{1}{3} & -\frac{2}{3}
\end{array}
$$

Then remains the problem of changing other remaining (four) elements not contained in the chosen row and column.

Form a rectangle with the pivot element and the element to be replaced such that the chosen row and the chosen column are two-sides

of the rectangle. If -2 of the second row is to be replaced, we consider the rectangle formed by values shown below:

$$\begin{array}{cc} -1 & (-2) \\ -3 & -2 \end{array}$$

Then the new value to replace -2 (shown within parentheses) is got by subtracting from the old value (-2), the quotient obtained on dividing the diagonal product not containing the pivot element, i.e.,

$$(-1) \times (-2)$$

by the pivot element (-3). Hence instead of

(*i*) -2 we will write $-2 - \dfrac{(-1)(-2)}{-3}$, i.e., $-\dfrac{4}{3}$

Similarly, instead of

(*ii*) 4000 we will write $4000 - \dfrac{(-1)(6000)}{-3}$, i.e., 2000

(*iii*) 1 ,, ,, $1 - \dfrac{(1 \cdot 5)(-2)}{-3}$, i.e., 0

(*iv*) 0 ,, ,, $0 - \dfrac{(1 \cdot 5)(6000)}{-3}$, i.e., 3000

Hence the new matrix becomes

$$\begin{array}{cccc} & & & y \\ z & 3000 & -\frac{1}{2} & 0 \\ & 2000 & \frac{1}{3} & -\dfrac{4}{3} \\ x & 2000 & -\frac{1}{3} & -\frac{2}{3} \end{array}$$

If we apply the rules again, we will now choose the last column and the second row. So, the new pivoting element will be $-\dfrac{4}{3}$ and the operations can be repeated as above to get a new matrix. This will leave x and y at the left and the slack-variables at the top. We will then have reached the optimal solution.

Non-linear Programming

In economics it is a common realisation that there are increasing or decreasing returns, including increasing and decreasing returns to scale (assuming at least one limited resource). Hence the programming-economists have gone further and are developing non-linear programming, wherein (*i*) the objective function to be maximised (or minimised), and (*ii*) the constraints may involve non-linear terms of the

variables such as x^2, sin x and 2^x. We may therefore conceive of maximising Z given by (say)

$$Z = 4x + (100 - \cdot 03y)\,y$$

where 4 is the fixed profit of a unit of x and $(100 - \cdot 03y)$ is the price of each unit of y according to market demand function. The maximisation is subject to

$$x^2 + y^2 \leq R_1$$
$$2x^2 + 3y^2 \leq R_2$$

where $x \nless 0$ and $y \nless 0$, R_2 and R_1 being given numbers.

The effect of the constraints in terms of graphical representation is easy to understand (Diagram 11.7). The first constraint will now appear as a curve and not as a straight line as shown by curve marked *I*. The feasible region will be that shaded by slant-lines. Similarly, the feasible region due to the second constraint will be the portion of the *II*-curve shaded by horizontal lines. Both constraints therefore reduce the feasible region to the doubly shaded area *OCRB*.

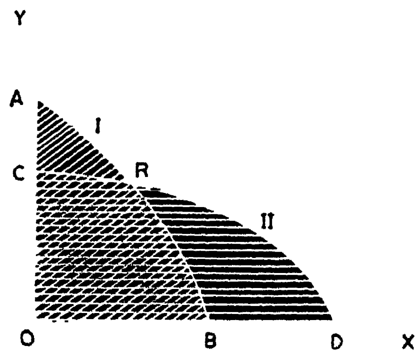


DIAGRAM 11.7

The maximising expression, which can be simplified as

$$4x + 100y - \cdot 03y^2$$

will now be represented not by a series of parallel straight lines but by a family of curves Z_1, Z_2, Z_3, . . . which are not necessarily parallel to each other.

Each Z-curve is really an iso-profit curve, i.e., each point on it stands for same profit Z.

We will now choose the position *R* (Diagram 11.8) which is reached by the highest Z-value curve. This point need not necessarily be one of the corners of the *CRB* boundary. Thus a great advantage of linear programming is lost. An optimal feasible solution cannot be found by merely examining the position at the corners. In fact, an optimal point need not imply a condition of tangency between the Z-curve and the boundary line: the optimal point *P* may well lie on a Z-curve at a point well within the feasible region *OCRB*.

Let it be stated without proof that where non-linear constraints or relations are involved, a solution by means of linear programming approximations may yield misleading results. Thus in the case of dimi-

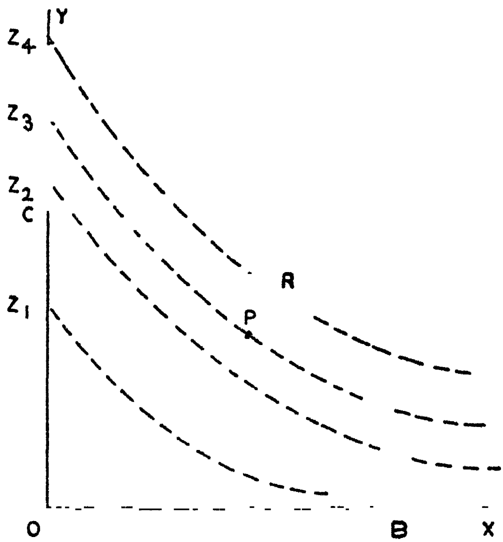


DIAGRAM 11.8

nishing returns it is likely to lead to too few activities; and in the case of increasing returns, to too many activities.

Dynamic Aspects and Linear Programming

In economics, commodities over time or place are treated as different commodities for purposes of dynamic and spatial analysis. Thus, if in an economy a single commodity is produced and part of it is used as input for production during the next period we can express the situation as a linear inequality.

Let a unit of the commodity lead to the production of α units of the same commodity in the next period. Let the output of this period be x_t, of the next period αx_t and let c_{t+1} out of αx_t be consumed leaving x_{t+1} to be used as input for further production. Then the total output of the next period can also be put down as

$$c_{t+1} + x_{t+1}$$

and obviously this cannot be greater than αx_t. Hence

$$c_{t+1} + x_{t+1} \leqslant \alpha x_t$$

becomes a production-inequality.

If the initial stock of the commodity was λ then

$$c_0 + x_0 \leqslant \lambda$$

and

$$c_1 + x_1 \leqslant \alpha x_0$$

Since $\quad x_0 \leqslant \lambda - c_0$

and hence $\quad \alpha x_0 \leqslant \alpha (\lambda - c_0)$

we can write

$$c_1 + x_1 \leqslant \alpha x_0 \leqslant \alpha (\lambda - c_0)$$

or

$$x_1 \leqslant \alpha (\lambda - c_0) - c_1$$

Since $c_2 + x_2 \leqslant \alpha x_1$, we can rewrite it as

$$c_2 + x_2 \leqslant \alpha [\alpha (\lambda - c_0) - c_1]$$

$$\leqslant \alpha^2 [\lambda - c_0] - \alpha c_1$$

or

$$x_2 \leqslant \alpha^2 (\lambda - c_0) - \alpha c_1 - c_2$$

If we generalise from here, we can write

$$x_t \leqslant \alpha^t (\lambda - c_0) - \alpha^{t-1} c_1 - \alpha^{t-2} c_2 \ldots\ldots - \alpha c_{t-1} - c_t$$

This is a linear relationship involving only the first powers of x and c's. If we want, we can fix the capital stock (x_t), know the initial output and prescribe consumption for all periods except one (say, that for period, $t-r$) and then we can use the relationship above to maximise c_{t-r}.

Similarly, if we had two processes, i.e., if what is not consumed were allocated to two production items in quantities x and y, we would write

$$c_0 + x_0 + y_0 \leqslant \lambda$$

$$c_1 + x_1 + y_1 \leqslant ax_0 + by_0$$

$$\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots$$

$$c_t + x_t + y_t \leqslant ax_{t-1} + by_{t-1}$$

where ax_0 and by_0 are respectively the outputs produced by using x_0 and y_0, i.e., a and b are output-coefficients.

Here again, given initial stock (λ) and final outputs and all (but one) consumption-quantities, we can try to use the linear programming device to maximise the remaining consumption item.

Existence Theorems

This is just to give an idea of how linear programming technique is being applied to dynamic problems and how it is adapted for non-linear programming. Since solving a linear programming problem involves a lot of expenditure of time for computations, to avoid useless attempts, Kuhn and Tucker have evolved what are called *existence theorems* which tell us whether a solution to a given problem exists. Of course they do not tell us how to go about finding that solution but only whether a solution exists. Even this is of great use.

PROGRAMMING AND REALITY

Finally, linear programming is based on some type of a model with an equilibrium and its use is of questionable utility in a world with ever-changing imperfect oligopolistic conditions. Besides, it must be remembered that the properties of a real world are too difficult to be represented in a model.

SOME PROBLEMS IN ECONOMICS

Before we go further, we give below some cases in economics where linear relations, either as equality or inequality, arise either in the form of (*i*) a budget or resource constraint, and (*ii*) short-period variations in costs (actually or as a first approximation or after logarithmic transformations); and hence in such cases, linear programming technique is usable:

1. If A has (say) Rs 100 and purchases k commodities in quantities $x_1, x_2, \ldots, x_k$ at prices $p_1, p_2, \ldots, p_k$, then

$$p_1x_1 + p_2x_2 + \ldots + p_kx_k \leqslant 100$$

2. If B has Rs 50,000 and is purchasing m inputs in quantities $y_1, y_2, \ldots, y_m$ at prices $u_1, u_2, \ldots, u_m$, then

$$u_1y_1 + u_2y_2 + \ldots + u_my_m \leqslant 50{,}000$$

3. If C has Rs 2,000 and is purchasing labour for use in 4 different ways at price p_1, then

$$p_1\,(l_1 + l_2 + l_3 + l_4) \leqslant 2000$$

4. If D has to incur a variable cost on labour and raw materials the marginal increase in cost will be approximately a linear function

$$p_l\,(\Delta l) + p_c\,(\Delta c)$$

where Δl and Δc indicate marginal increases in labour and raw materials for an increase Δx in output, and p's are prices.

5. If a producer E experiences a Cobb-Douglas production function given by

$$x = a\,\lambda^t\,l^\alpha\,c^\beta$$

where $x =$ output, $t =$ time, $l =$ labour, $c =$ capital (a, λ, α and β being constants or parameters), then on taking logarithms we get

$$\log x = \log a + t \log \lambda + \alpha \log l + \beta \log c$$

On putting $\log x = x'$, $\log l = l'$, $\log c = c'$ and replacing the constant $\log a$ and $\log \lambda$ by A and B respectively we get

$$x' = A + Bt + \alpha l' + \beta c'$$

which is a linear relation expressing x' in terms of t, l' and c'.

6. A producer F has three machines which can be worked for times t_1, t_2 and t_3 respectively to produce two commodities A or B. The rates of production for the first machine are respectively (*i*) r_{11} and r_{12} for A and B. For the other machines the rates are similarly (*ii*) r_{21} and r_{22} and (*iii*) r_{31} and r_{32}. When sold, the commodities A and B yield a profit of α and β per unit of sale. The situation is then representable as follows where X_A and X_B are outputs of A and B:

$$\text{Profit} = \alpha X_A + \beta X_B$$

$$\frac{X_A}{r_{11}} + \frac{X_B}{r_{12}} \leqslant t_1$$

$$\frac{X_A}{r_{21}} + \frac{X_B}{r_{22}} \leqslant t_2$$

$$\frac{X_A}{r_{31}} + \frac{X_B}{r_{32}} \leqslant t_3$$

These relations are all linear. The profit expression has to be maximised subject to three linear constraints or conditions of inequality.

7. A trader G has three godowns A, B and C and has to supply orders to customers located at four places P_1, P_2, P_3 and P_4. The stocks are 30, 40 and 53 units at A, B and C. The demands are for 22, 35, 25 and 41 units from P_1, P_2, P_3 and P_4. The transport charges from A, B and C to P_1, P_2, P_3 and P_4 are as follows:

	P_1	P_2	P_3	P_4
A	23	27	16	18
B	12	17	20	51
C	22	28	12	32

The problem is to allocate the orders so as to minimise the total transport cost of distribution.

It will be seen that the (linear) total of assignments of order to each godown cannot exceed the stocks there. The total transport cost will also be a *linear* function of supplies and transport rates and this has to be minimised.

8. A producer H can pack different combinations of a number of commodities in a finite number (or, variety) of packets subject to the condition that each packet meets a particular minimum requirement or standard. He has to decide how he should prepare the mixture to minimise his cost of production. This also involves the minimisation of cost subject to linear constraints.

9. K has to maximise the production of machines in India and Japan together subject to a prescribed total production of food T_f, it being given in what ratio food and machine can be produced in each country out of given resources R_I and R_J in the countries respectively.

If in India a unit of resource (measured in food units) produces 1/5 unit of machine or 1 unit of food and if the Indian outputs of food and machine be x_f and x_m, then

$$x_f + 5x_m \leqslant R_I$$

since total resources used up in India cannot exceed R_I.

Similarly, if in Japan a unit of resource (also measured in food units) can lead to a unit of food or $\frac{1}{2}$ unit of machine we may write

$$y_f + 2y_m \leqslant R_J$$

where y_f and y_m are outputs of food and machine in Japan.

We have therefore to maximise Z, where

$$\left.\begin{aligned} Z &= x_m + y_m \\ \text{subject to} \quad x_f + y_f &= T_f \\ x_f + 5x_m &\leqslant R_I \\ y_f + 2y_m &\leqslant R_J \end{aligned}\right\}$$

Of course x's and y's are all non-negative values.

This is a linear programming problem: we can easily express the equality as an inequality

$$x_f + y_f \geqslant T_f$$

if there is no harm in overshooting the food target.

10. If India wants to maximise her national product in food units when it has the option of producing food or machines as given in illustration '9' above and if the price of machines in terms of food units be somewhere between 5 and 1 (say, at 3), then the problem of what it will produce can be posed as a linear programming problem. We would maximise Z given by

$$Z = 3x_m + x_f$$

subject to the production-possibility curve

$$5x_m + x_f = R_I$$

The linear programming method has also been used in the following applied economic problems:

1. *Machine allocation:* A number of orders have to be allocated to various machine groups to ensure minimum cost or least "over-all running time".
2. *Inventory management:* When sales show seasonal fluctuations, an inventory programme is to be set up to reduce storage to the minimum.
3. *Rail or truck movements:* Given the table of distances from each place to others for a number of places, and given a required amount of movements, we would like to minimise the total mileage-run.
4. *Market research:* Chain stores have to choose sites from among a number (sometimes several hundred) of sites subject to maximisation of the weighted total of measures of each site.

5. *Advertising evaluation:* Lakhs of rupees are spent on advertisements and it is a problem to choose media for advertisements to maximise sales.

LINEAR PROGRAMMING AND THE THEORY OF THE FIRM*

We know that marginal analysis assumes (1) a continuous production function, (2) any degree of substitution, and (3) availability of factors without specified limit. But in a firm decisions are rather short-run decisions, resources are limited (say, in the form of cash-credit, land, buildings etc.) and decisions consist in choosing between a finite number of processes, each of which is getting more complex and less flexible than marginal analysis assumes.

Linear programming comes in as a handy alternative, or supplement for producing answers to practical questions of behaviour of a firm. We assume that (1) there are a small number of processes, each with its inflexible technological input-baskets and that any process can be used at any level and in combination with any other process. To keep calculations to manageable limits, each process is assumed to be subject to constant returns to scale. Further, for each process a unit-level of activity is assumed. It is generally with reference to the quantity or (usually) value of the output: sometimes it may be with reference to, say, the use of certain manpower. Thus it may be given that for producing Rs 1,000 worth of the same good we need man-hours and machine-hours as follows:

	Process No. 1	*Process No. 2*	*Process No. 3*
Man-hours	50	25	80
Machine-hours	5	50	60

We may represent each of these unit level process on a graph (Diagram 11.9) by the points P_1,P_2 and P_3. The straight line OA_1 can then be taken to indicate the different levels of production by process No. 1: it is the locus of positions of points like P_1.

Similarly, OA_2 and OA_3 can be interpreted as loci of levels of production by the other two processes.

The feasible region of production at unit level, i.e., to produce Rs 1,000 worth of goods, is the triangle $P_1P_2P_3$. Any point within the triangle space can be interpreted as representing a combination of the processes. But, as is evident from the graph, the side P_1P_2 of the triangle

*A useful exposition of analytical and graphical comparison of marginal analysis and programming is given in Chapter 4 of *Linear Programming and the Theory of the Firm* by Boulding and Spivy (1961).

is apparently nearest to the origin and *hence* (?) represents the most economical use of the given resources. Therefore it seems reasonable to ignore the triangle and accept P_1P_2 as the line-space of efficient feasible solutions at unit level.

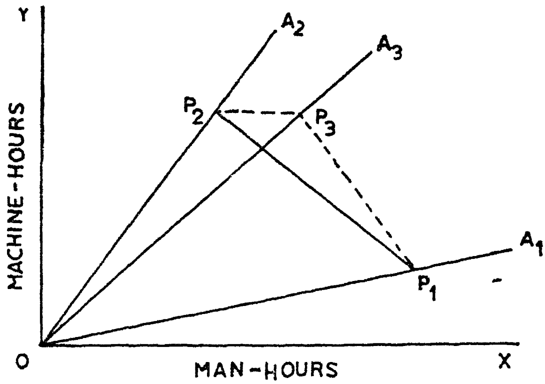


DIAGRAM 11.9

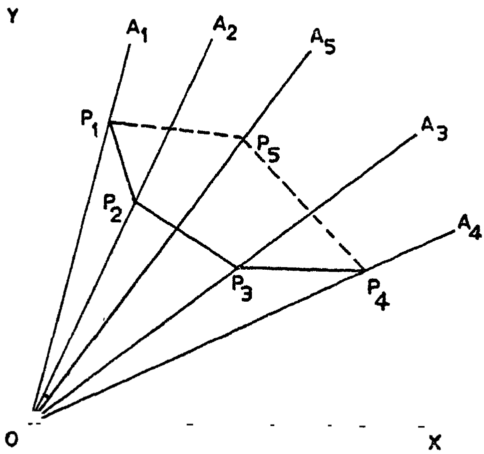


DIAGRAM 11.10

In general, we may conceive of many (say, 5) processes indicated by OA_1, OA_2, OA_3, OA_4 and OA_5, with unit level of activity at P_1, P_2, P_3, P_4 and P_5. (Diagram 11.10) Then the feasible region for unit production will be the space $P_1P_2P_3P_4P_5$ and the efficient feasible boundary space will be $P_1P_2P_3P_4$ only as this is most convex to the origin.

The boundary-line $P_1P_2P_3P_4$ is akin to an iso-product curve. It is not a smooth curve because all levels of combination of processes are not feasible.

It may be submitted that if we draw three (or more) axes, i.e., if we had three (or more) factors to reckon with, it may be difficult to determine which is the most convex boundary-space. For that purpose the simplex-method is more mechanical and helpful than the graphical method.

However, like a family of iso-product curves, we may conceive of a family of iso-activity boundary space.

We may then, graphically speaking, *either* choose the B curve which makes full use of the given resources for maximum production (*irrespective of needs*) *or* choose the B curve nearest to a maximum (or minimum) value of objective function, in which case the decision will be (to an extent) in keeping with the demand side at given prices.

Thus if Q (see Diagram 11.11) indicates the available resources OM of one factor and ON of another, we shall choose the B_3 curve and produce goods through a combination of A_2 and A_3 processes.

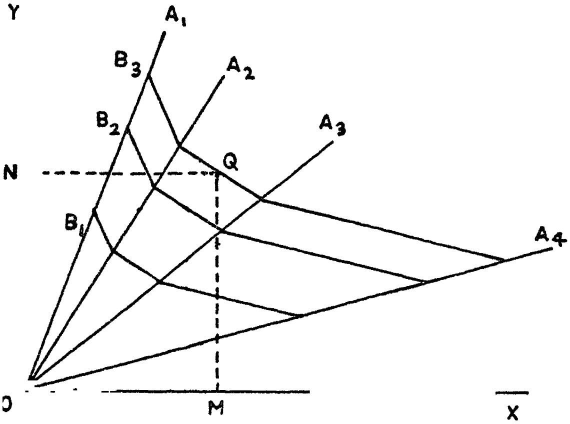


DIAGRAM 11.11

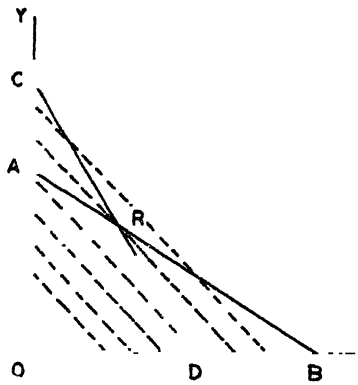


DIAGRAM 11.12

Alternatively, appropriate to the level (λ) of activities of the two processes A_2 and A_3 we may draw the constraint-curves AB and CD (Diagram 11.12) and, with the help of the family (Z) of objective function, choose the maximum point (R).

Thus, let the processes be as detailed below:

Resource	*Process No. 1*	*Process No. 2*	*Process No. 3*	*Resources available*
Labour-hours	50	25	80	50
Machine-hours	5	50	60	30
Produce	1	1	1	

Then, if the level of activities selected ultimately be λ_1, λ_2, and λ_3, the total produce is the total of produce from each process, i.e.,

$$Z = 1.\lambda_1 + 1.\lambda_2 + 1.\lambda_3$$

subject to the constraints

$$50\lambda_1 + 25\lambda_2 + 80\lambda_3 \leqslant 50$$
$$5\lambda_1 + 50\lambda_2 + 60\lambda_3 \leqslant 30$$
$$\lambda_1 \geqslant 0; \lambda_2 \geqslant 0; \lambda_3 \geqslant 0$$

The last condition is added because no level of activity can be negative.

If, on the basis of the graphical representation, we reject the third process, the problem would reduce to

$$Z = \lambda_1 + \lambda_2$$
$$50\lambda_1 + 25\lambda_2 \leqslant 50$$
$$5\lambda_1 + 50\lambda_2 \leqslant 30$$
$$\lambda_1 \geqslant 0; \lambda_2 \geqslant 0$$

and the solution can be found in the same way as in the consumption or production example given earlier.

It may be mentioned that no consideration is given here to the timings of labour-hours and machine-hours. We assume that these resources are available whenever the need arises during the process of production. Suppose a process of production takes five months to complete (e.g., in agriculture), then the resources (except land) are not required throughout the five months uniformly. Linear programming does not take account of what happens to men and machines when these are not required for (say) agriculture.

Besides, to the extent that an indivisible factor is involved, which admits of greater or less use, we will find that, with varying level of activity, the factor-coefficients (such as labour-hour 50, machine-hour 25 in Process No. 1) will change relative to land. But linear programming does not take account of such increasing or decreasing intensity of use of a fixed factor. Researches are however being made to extend the scope of linear programming to take account of cases of (*a*) a mixture of fixed and hired factors, (*b*) demand for products in fixed proportions, e.g., razor and blade or tobacco and smoking pipe, and (*c*) case of two or more producers who operate separately but who may cooperate for a better joint result.

Linear Programming and Planning

How linear programming can be used for investment planning may be just illustrated. Suppose the government has two types of power stations and wants to construct them so as to increase the production capacity by C such that the stations meet a maximum (or peak) demand of M_x and guarantee a minimum capacity M_i. Let the technological details be as follows:

	Type of Power Station	
	I	*II*
Guaranteed capacity	1	1
Peak capacity (MW)	1·1	1·2
Yearly output (GWH)	7	1·3
Construction cost	K_1	K_2
Operating cost	O_1	O_2

If the stations are to be of capacity x_1 and x_2, we can say that

$$x_1 \geqslant 0;\ x_2 \geqslant 0, \text{ and}$$

combined guaranteed capacity $= x_1 + x_2$ which is $\geqslant M_i$
combined peak capacity $= 1{\cdot}1x_1 + 1{\cdot}2x_2$ which is $\geqslant M_x$
combined output $= 7x_1 + 1{\cdot}3x_2$ which is $\geqslant C$
combined construction cost $= K_1x_1 + K_2x_2$
combined operating cost $= O_1x_1 + O_2x_2$

Subject to the constraints above, we may minimise any one of the following to ensure efficiency of investment:

$$(i)\ Z = K_1x_1 + K_2x_2$$
$$(ii)\ Z = O_1x_1 + O_2x_2$$
$$(iii)\ Z = (K_1x_1 + K_2x_2) + 1/r(O_1x_1 + O_2x_2)$$

where r is the rate of discount.

Linear Programming and Consumption*

In applying linear programming to consumption problems with an eye on empirical utility, it is a weak assumption that goods yield constant

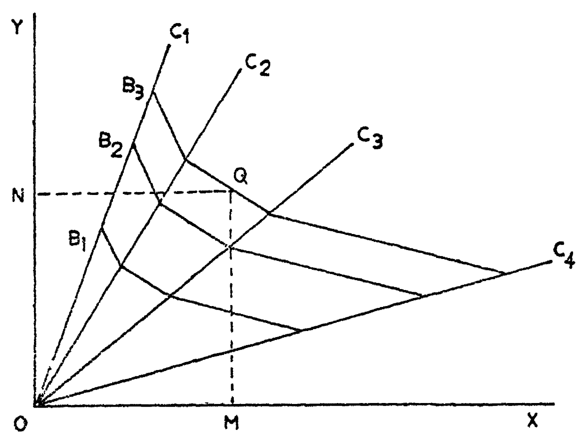


Diagram 11.13

* Vide R.G.D. Allen: *Mathematical Economics*, Sec. 19.6.

returns of utility to scale. Besides, some goods are competitive, some complementary. There is also the question of interpersonal comparison of utility. Hence, if we want to apply linear programming to the behaviour of consumers in a market, the necessary assumptions would seem to take the application too far from reality. If we are prepared to assume the existence of an indifference map for the products of a firm (or firms) for their consumers and conceive of a consumption activity as an activity with consumer goods in a fixed proportion as inputs, we may well draw a system of boundary-lines and, as in the case of production, we may bid to reach an optimum solution at *Q*. (Diagram 11.13).

Linear Programming and Welfare Economics*

It is a well known theorem in welfare analysis that a long-run competitive equilibrium yields an optimal allocation of resources. But it is true only under certain fairly restrictive assumptions, and for commodities which are saleable without loss and which are not free. Whether or not the theorem applies will depend on costs of production and pattern of demands into which commodities fall in the above-mentioned category of 'commodities'. The theorem does not rule out a solution with negative consumption and negative prices. But the linear programming method can be pressed into service for we assume:

(*i*) that no prices and quantities are negative (non-negativity condition),

(*ii*) that production cannot exceed the levels possible with available resources, and

(*iii*) that average cost of production (*a*) exceeds the price at which non-saleable goods can be sold, and (*b*) is equal to the price for all other goods.

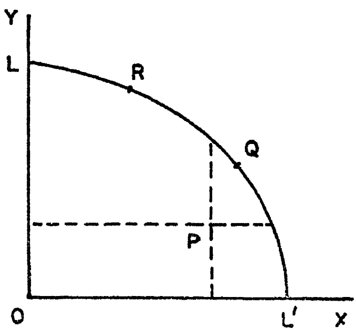


Diagram 11.14

If *OX* and *OY* (Diagram 11.14) are used to indicate the outputs of two commodities and if for each of successive chosen values of one commodity we find (using programming techniques) the maximum output of the second commodity, the locus of the two values will be a curve in the *XY*-plane. We call it a *transformation locus LL'*.

Each point on the locus *LL'* is optimal or efficient. The area *OLL'* is a feasible region of production. A point *P* inside it

* Vide W.J. Baumol: *Economic Theory and Operations Analysis*, Chap. 16.

indicates a feasible production-basket. All points on the locus LL' lying in the North-East quadrant with P as origin are more efficient than P, but it is difficult to say whether R (also a point on LL') is more efficient than P. However, we know that competitive equilibrium involves a maximum value of output at equilibrium (fixed) prices. Now at fixed prices, an inefficient output cannot mean a maximum money value, for, through greater efficiency, output of at least one commodity can be increased without decreasing the output of any other commodity. Hence, maximum value corresponds to the most efficient output-basket. Hence, competitive equilibrium must necessarily correspond to efficient production.

The converse, viz., that every efficient output-basket is a competitive output-basket has also similarly been proved.

Where indivisibilities and increasing returns complicate a situation, integer programming analysis can be pressed into service.

CHAPTER 12

Aggregation

Nature of Aggregation

AGGREGATION is an act of collecting or adding together. An individual aggregates the various satisfactions from different consumptions to form the idea of his 'total satisfaction'. A trader aggregates the sales proceeds from various goods to get his 'total sales'. An individual adds up quantities of satisfaction, but a trader may add up the different quantities of goods sold or their values in money terms. But an addition of different quantities in different units is impossible and meaningless. Two tumblers, three kilograms of phenyl and two metres of ribbon cannot be added, but if they were sold for two rupees, twelve rupees and one rupee respectively, we can add these and put in total sales as fifteen rupees.

Similarly, aggregation is necessary when we think of the market demand for a commodity. In that case one may total the sales of that commodity made by the different sellers either in physical quantity or in terms of money-value. Thus, at a price of Rs. 1·37 per kilo, we may find out and add the quantities of sugar sold by different dealers in Allahabad in order to find the sale of (or, demand for) sugar in that city. But if we want to study sales of sugar, wheat, matches and soap, all taken together, then the aggregation problem becomes difficult and awkward. We can, then, at best, add their sales in money terms. However, if we study sales over a period of time, we may convert the series of sales of each commodity into a series of index numbers and then average the index

numbers of sales by quantity of sugar, wheat, matches and soap. We can do so for sales in money terms also.

In economics a study is sometimes made at the micro level and sometimes at the macro level. The former falls under microeconomics; the latter, under macroeconomics. In a society we have many goods and markets. There are many interrelations. In our study of these interrelations we may increase the number of variables and relations. We may write out Walrasian equations. Alternatively, we may reduce the number of variables and relations as we do in input-output analysis where some aggregation will be necessary. At the extreme, we may think only in terms of income, consumption, savings and investment. In that case we will need (say) (1) aggregation of all incomes, (2) aggregation of all types of consumption, (3) aggregation of all types of investments in building, machines, trucks and raw materials.

In economics aggregation means (*i*) sometimes addition of quantities, values, flows and stocks, and (*ii*) sometimes determination of a representative value as average price and average rate as also price-level and production level.

Why Aggregation

At this stage let us be clear about why we undertake or should undertake aggregation. To an extent, the growing, complex, large-scale economic activities make it essential that aggregation of some type be practised to enable a board of directors to take decisions even at the micro level. Aggregation is necessary if we are not to be lost in a wood of details of innumerable goods and services, individuals and firms, localities, markets and regions.

In order that we may not fail to see the wood for the trees we do (and we should) think in terms of aggregate demand, aggregate supply, general price level, national income, national consumption, national investment, and so on. This will enable us to analyse the activities of a society, not at the level of individuals, but at that of all individuals taken together. This will keep down the number of variables and relations. Thus, we may then write:

$$Y = C + I$$
$$C = \alpha Y + \beta$$
$$I = gY + h$$

Thus, for such simplified analysis aggregation is inevitable.

With the advent of central planning for economic development, such aggregation is indispensable for policy decisions, action and review.

How Aggregation

Before aggregation is attempted, a decision must be taken on units of time, space and measurement. A specific period of time (say a week, a month, a season or a year) must be selected and assumed to be a unit of time for purposes of aggregation. We aggregate the production of different commodities during a week, a month or a year. We aggregate the purchases made during a week, a month or a year. Similarly, we find the average number of employees per year and average price prevailing during a year.

Again, a unit of space is also necessary for aggregation. Thus we may aggregate production by all producers in the U.S.A. or by all producers in India.

Finally, aggregation can be undertaken only after expressing such measurements in terms of a common denominator or unit. This is invariably the 'money of account.' Thus we study national income, consumption, savings and investment in India in terms of rupees.

Thereafter, aggregation is *prima facie* easy. One has to aggregate or add the values. In certain cases, e.g., prices and employment, one has to take some sort of an average to arrive at a (aggregate) general price-level and employment.

Problems of Aggregation

Three problems arise in regard to aggregation:

(*i*) What groups are to be formed and which items are to be included in the different groups for which aggregation is to be practised?

(*ii*) At what prices are the items to be evaluated for purposes of aggregation?

(*iii*) If aggregation is to be used for studying behaviour over time, such as from month to month, season to season, or year to year, how shall we eliminate the effects of changing prices so as to be able to study the actual physical variations?

1. *Grouping:* This mainly depends on the purpose of the study and the importance attached to particular modes of analysis. We can understand this easily by taking illustrations. When we aggretate demand, we do not aggregate all demands arising in a country. Instead, we undertake separate aggregation for consumption-demand and investment-demand. Consumers' goods enter the first group and producers goods are reckoned in the second group. We, therefore, get C and I of the model

referred to above. For many a study, the buyers are divided into four groups:

1. Consumers
2. Firms
3. Government
4. Foreign customer

Similarly, supply may be aggregated into internal supply and external supply (i.e., imports).

For purposes of input-output analysis (See Chapter XIII) we sometimes divide economic activities into:

(1) agriculture
(2) industries
(3) final demand
(4) stocks

for flow of output, and into

(1) agriculture
(2) industries
(3) government
(4) primary producers
(5) imports

for purposes of inputs.

Aggregation has then to be practised for these groups. The boundary lines of the groups are always vague. Thus, how many cars should fall under consumers' goods and how many under investment goods? How much coal should be classified under consumption and how much under investment? One can easily think of other such examples.

However, the purpose of study influences the groups under which aggregation is done. If the purpose is to study the existence of equilibrium, a static analysis may be relevant and we may ignore consideration of stocks of capital in aggregating output and demand. Since aggregation is meant to replace a detailed theoretical model by a simplified model, a decision about groups must depend on the theoretical model to be followed in spirit. In practice, aggregation may have to be changed or readjusted for purposes of application.

In general, it may be said that groups should be such that

(1) commodities falling in a group are perfectly substitutable in their uses;

(2) there is no possibility of substituting the goods of one gorup for those of another; and

(3) the technique of production is the same within the group.

These criteria are not mutually exclusive in one sense. Thus, two commodities may fall in one group yet be separable from the point of technique of production, e.g., yarn and cloth: or, they may be substitutes and yet fall in different groups, e.g., (*a*) coal, petroleum and

electricity, or (*b*) cotton, silk and wool. If we keep spinning (yarn) and weaving (cloth) in separate groups there will be difficulty in properly separating data from mills which undertake both spinning and weaving.

At times, aggregation may be practised for a policy decision (say, about total textile production) and then a subdivision or disaggregation may be made for outputs of cotton, wool and silk textiles. This sometimes involves a difficulty in actual mathematical treatment. Thus, total textile production may turn out to be x and when it is subdivided mathematically into

x_c — for cotton textiles
x_w — for woollen textiles
x_s — for silken textiles

one of the three may turn out to be negative, while implicitly no x can be negative. Such a possibility exists more in the case of aggregation and disaggregation of intermediate-goods sectors in input-output analysis.

Hence we may finally say that, for a decision on grouping, not only do we need to have full knowledge of the economic field under examination and the purpose of study but we also need to be prepared to use our commonsense and intuition to finalise grouping.

2. *Pricing:* Since we aggregate values, the question of prices at which physical quantities are to be valued arises. Valuation may be at *factor cost* or at *market price*. The former refers to prices paid for factors of production: thus, if B produces 100 units for Rs 1,000, the factor cost is Rs 1,000, even though when sold in the market these 100 units secure Rs 1,200.

In input-output analysis this question is more pertinent. The intermediate flows are generally valued at factor cost while the final demand is valued at market price.

Again, when we are aggregating all cotton textiles, as reported in the statistical returns of textile mills and handloom weavers, we may have to convert the physical quantities into 'values' by multiplying them by *appropriate* prices. What prices shall we then adopt?

Where we must evaluate things at one common price for the sake of simplification, we must come to a decision as to which average of a whole series of relevant prices we are to take. The average may be an arithmetic average, a geometric average, a mode or a weighted average. If we take a weighted average, what weights shall be adopted?

Many a time, an index number of prices is used for the purpose of evaluation and the question then is, "Which index number shall be used?"

To give an illustration of the difficulty of pricing, let us consider the following case. There are two commodities C_1 and C_2 produced in

quantity 15 and 9 in 1965 and, 30 and 1·5 in 1966. The corresponding prices were (1) 9 and 20, and (2) 10 and 16. Then the valuation for 1965 and 1966 are as follows:

$$1965: \quad 15(9) + 9(20) = 315$$
$$1966: \quad 30(10) + 1{\cdot}5(16) = 324$$

If we decide to use the same prices in both years then we shall have the following results:

	Year	*Valuation*	
At 1965 prices	1965	15(9) + 9(20)	= 315
	1966	30(9) + 1·5(20)	= 300
At 1966 prices	1965	15(10) + 9(16)	= 294
	1966	30(10) + 1·5(16)	= 324

Thus we find that the production (by value) goes up according to both current prices (315 to 324) and 1966-prices (294 to 324) but goes down according to 1965-prices (315 to 300). Should we use current prices, 1965-prices or 1966-prices for evaluation of the quantities produced in both years? Or, should we use some index of prices?

3. *Study over Time:* In practice, economic characteristics and relationships among them are studied over time — say, over a number of years. Then, many a time, it becomes desirable to eliminate the effect of changing prices. For this, the values for different periods are converted into values at constant prices by dividing the various values by an index number of prices.

Thus if we have the following aggregated values

Year	1963	1964	1965	1966
Income	12	12·8	13·3	15·3
Consumption	9	9·6	10·5	14·0
Price-Index (1960-61 = 100)	105	120	140	180

we may divide each value of income and consumption by the price-indices and get the values for income at the prices prevailing in 1960-61:

Year	1963	1964	1965	1966
Income	$(12)\,\frac{100}{105}$ $= 11{\cdot}4$	$(12{\cdot}8)\,\frac{100}{120}$ $= 10{\cdot}7$	$(13{\cdot}3)\,\frac{100}{140}$ $= 9{\cdot}5$	$(15{\cdot}3)\,\frac{100}{180}$ $= 8{\cdot}5$

These are deflated values of income at 1960-61 prices.

Similarly, we may obtain the deflated values of consumption:

Consumption	8·6	8·0	7·5	7·8

Here one has to consider whether the same series of index numbers

should be used for deflating income and consumption values. Income is made up of consumer goods and producer goods, and the prices of producer goods may not vary in the same manner as those of consumer goods.

Besides, how should the index number of prices be constructed? In fact, as we have seen already, the question of index number construction arises even in the study of levels of production. However, in regard to price-levels over time one has to study the value of a constant basket of commodities and services at different dates:

$$\text{Price level} = \frac{\text{Money value of a basket of commodities}}{\text{Total physical quantity in the basket of commodities}}$$

Usually the price-levels appear as indices of prices. Thus we may have:

(1) $\text{Price level} = \dfrac{\text{Index of value}}{\text{Index of quantity}}$

(2) $\text{Price level index} = \dfrac{\text{Value of a basket in current year}}{\text{Value of the same basket in a previous year}}$

Thus, in our two-commodity example, we may compare

(1) the value of a basket of 15 and 9 units in 1966 with its value in 1965

$$\begin{aligned}\text{Price index} &= \frac{15(10) + 9(16)}{15(9) + 9(20)}\\ &= \frac{294}{315}\\ &= 0{\cdot}93\end{aligned}$$

(2) the value of a basket of 30 and 1·5 units in 1966 with its value in 1965:

$$\begin{aligned}\text{Price index} &= \frac{30(10) + 1{\cdot}5(16)}{30(9) + 1{\cdot}5(20)}\\ &= \frac{324}{300}\\ &= 1{\cdot}08\end{aligned}$$

(3) the values of some other representative basket of commodities. In technical language, the weights for averaging the prices may be chosen in a variety of ways:

$$\begin{aligned}\text{Price index} &= \frac{\Sigma wp_1/\Sigma w}{\Sigma wp_0/\Sigma w} = \frac{\text{weighted average of prices in 1966}}{\text{weighted average of prices in 1965}}\\ &= \frac{\Sigma wp_1}{\Sigma wp_0} = \frac{\text{value of basket } w \text{ at prices in 1966}}{\text{value of the basket } w \text{ at 1965 prices}}\end{aligned}$$

And should the weighted average be arithmetic average (as above) or geometric average or some other average? The statistician discusses this in connection with the construction of an ideal index number of prices; but the criteria of the ideal are not yet universally agreed upon.

Convenience of calculation and ease of understanding has led to the use of the weighted arithmetic average.

Which prices to choose? is another problem. Should we form an index of wholesale prices or retail prices or harvest prices of foodgrains in order to form an index of foodgrains prices?

INDEX NUMBER PROBLEM AND WELFARE OVER TIME

If welfare is taken to be proportional to the value of goods produced, a study of welfare over time can be made by studying changes in the aggregate value over time. How can we measure aggregate value? In practice, the answer is: Form an index of aggregate value.

Aggregate value index = Index of quantities × Index of prices. But there may be several different indices of quantities and several different indices of prices. Which ones should be chosen for the purpose of an aggregate value index is a vexing problem. *In practice,* these indices are weighted averages and if the weights for quantity-index are this year's prices then the weights for the price-index are the last (base) year's quantities or vice versa:

$$(i) \quad \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} \cdot \frac{\Sigma q_0 p_1}{\Sigma q_0 p_0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}$$

$$(ii) \quad \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \cdot \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0}$$

as these lead to consistent results.

THE AGGREGATES CALLED NATIONAL INCOME AND NATIONAL ACCOUNTS

One may also refer, in connection with aggregates, to the vagueness and limitations associated with the concepts of national income and national accounts. An idea of these accounts was given in the chapter on Social Accounting. We will therefore only refer to a few criticisms on the lines of Boulding:

(1) What is termed household consumption is really a mixture of current consumption plus investment in durable consumer goods. The greater the element of durable consumer goods, the greater should be the effect on productivity but the present practice ignores this aspect.

(2) Family labour is ignored. The national income figures have no place for the value of household labour put in by the housewife.

(3) Government expenditure may be looked upon partly as a consumption item and partly as an investment item. But whether a parti-

cular expense is one or the other may be controversial. Thus, even defence expenditure may be termed as a consumption expenditure incurred to fill the consumer with the satisfaction at the fact that his country is a mighty country. Alternatively, it may be looked upon as an inevitable overhead cost of the continued peaceful operation of the production-machinery. National accounts leave this vague.

(4) Finally, difficulties of data collection have made it inevitable to adopt compromise methods in such respects as allocation of government taxes and transfers to different functional heads.

CHAPTER 13

Input-Output Analysis

INTRODUCTORY

INPUT-OUTPUT analysis is a tool of economic analysis which enables us (*i*) to take account of interdependence of economic activities in the study of general economic equilibrium, (*ii*) to test the internal consistency of a programme or plan with a view to locate (and remove) bottlenecks, and (*iii*) to predict the gross outputs needed in the various sectors of economic activities in order to achieve an objective, which may be (say) any one of the following:

(*i*) maximum profit
(*ii*) maximum rate of growth of national income
(*iii*) optimal employment
(*iv*) maximum foreign exchange earnings
(*v*) maximum savings

Input-output analysis is also called

(*a*) analysis of inter-industry (or inter-sector) flows (or deliveries)
(*b*) analysis of inter-industry relations
(*c*) multi-sector analysis
(*d*) activity analysis

Input-output analysis has been used in connection with military and wartime problems also. In fact, much of it was developed during World War II.

HISTORY

Historically speaking, the input-output analysis concept can be seen in the *Tableau Economique* of Quesney and may be said to be a modern attempt to give a practical shape to the general equilibrium analysis of Walras, though this is controversial in view of the following arguments:

(*a*) It does not have behind it any hypothesis as to rational behaviour of producers and consumers.

(*b*) It is indifferent to the maximisation principle in as much as any objective can be behind it.

(*c*) It does not contain the mechanism of price-adjustment.

(*d*) It does not analyse what is called 'final demand' in input-output analysis. It begins and ends with production analysis. It does not introduce utility functions, and, hence, input-output analysis is not an analysis of an equilibrium system.

(*e*) It assumes the constancy of the quantity of a factor (or input) required per unit of produce (or output).

However, the reader should consider these arguments after he has learnt the details of input-output analysis. Reverting to history, this modern development may be attributed to discussions of national economic balance in the U.S.S.R. in the twenties. Leontief himself may have got his idea from there: he no doubt developed and gave it the form in which we shall see it below. His first input-output table related to a fixed period of time: with that period as a unit of time, the table presents a static picture and the analysis becomes static. Also, his analysis was in terms of variable inputs corresponding to variable costs rather than fixed inputs (or capital inputs!) corresponding to fixed costs. Present-day endeavours are to make input-output analysis dynamic and to make it separately for current (or variable) inputs and capital (or fixed) inputs.

PRESENT DAY INTEREST IN SUCH ANALYSIS

Today, there is widespread interest in input-output analysis all over the world. Almost every industrialised country has built up input-output tables. The developing countries in Latin America, Africa and Asia (including India) are rapidly developing this type of analysis. Even East-European countries are exploring it, though perhaps they may be finding (as in the U.S.S.R.) a system of *inter-division balances* superior.

Input-output analysis is no longer of only academic and research interest. It plays an important part in preparing an economic policy—

for economic forecasting in developed countries and for economic planning (or programming) in developing countries.

As in the case of national accounting, in regard to input-output tabulation also, an attempt is being made to standardise and normalise terms and concepts so as to ensure easier comparability among nations. Such an attempt has its drawbacks also, one of them being 'hindering of further developments', and so there is great interest in the experiments (in this regard) being carried out in the six countries of the European Community.

Today, economic activities may, in general, be said to be characterised by the following features:

(*a*) On the demand side, output depends, *inter-alia*, on
- (*i*) final demand by consumers (or final user)
- (*ii*) demand by government
- (*iii*) demand by foreigners (exports)
- (*iv*) demand by intermediate users for current use or for building up inventory (or stock)
- (*v*) prices;

(*b*) On the supply side, output is *inter-alia* a function of
- (*i*) supply of (internal) inputs — raw materials, labour, machinery, etc.
- (*ii*) supply of imports (external inputs)
- (*iii*) taxes and subsidies
- (*iv*) wages and other input-prices
- (*v*) technological requirements in respect of durable and non-durable inputs

Hence it seems apt that, in applied economics, the emphasis which was laid on studying and forecasting business cycles during the twenties and on market analysis during the thirties has now shifted to programming and input-output analysis. Planned economic development, whether on the national or the regional level, has necessitated a study of:

(*i*) requirements of gross outputs (including outputs of intermediate goods) for given target levels of final demand by consumers;

(*ii*) extent of dependence of performance in each production-sector on that in other sectors;

(*iii*) effects of changes in government purchases, exports and investment demands on necessary outputs in different sectors;

(*iv*) effect of wage-changes and price-changes; and

(*v*) effect of a change in the volume of imports, level of employment and indirect taxes.

FORMS OF INPUT-OUTPUT TABLE—SKELETON

It is, therefore, not enough to study national output as a sum of the quotients got by multiplying labour employed in different sectors by the labour-productivity in corresponding sectors. Labour productivity cannot be taken to be reflective of the role of other factors of production. Hence we conceive of the structure of economic activities as illustrated in the following table:

From \ To	E_1	E_2	E_3	E_4	*Total output*
E_1	x_{11}	x_{12}	x_{13}	x_{14}	x_1
E_2	x_{21}	...	...	...	x_2
E_3	x_{31}	...	...	...	x_3
E_4	x_{41}	...	...	...	x_4

This is a miniature input-output table. For its purpose, E_1, E_2, E_3 and E_4 are four activities. The output from each activity *flows to* activities E_1, E_2, E_3 and E_4. Thus the flow of E_1's output (x_1) is comprised of x_{11}, x_{12}, x_{13} and x_{14} which go to the four activities respectively. Similarly x_{11}, x_{21}, x_{31} and x_{41} are inflows or inputs used up in economic activity E_1. Each row, therefore, indicates the distribution of output of an economic activity and each column shows up the quantities of inputs used up for an output. So, inputs lead to outputs. Thus

$$x_{11}, x_{21}, x_{31}, x_{41} \rightarrow x_1$$

On the left hand side we have inputs: the right hand side is the output relating to E_1. The quantity x_{11} means that the part x_{11} of the output x_1 of activity E_1 is used up within itself.

We give below, by way of illustration, a number of input-output tables:

FIRST FORM (*£millions*)

Paid by \ Received by		*Industries* A	*Industries* B	*Industries* C	*Consumers*	*Totals*
Industries	A	20	40	60	100	220
	B	60	10	50	150	270
	C	60	110	80	250	500
Consumers		80	110	310	—	
	TOTAL	220	270	500	500	

Here there are three industry-groups. The fourth group is that of consumers who supply their services and get payments as shown in the row against consumers. Consumers get 80 + 110 + 310 i.e., £500 millions and spend 100 + 150 + 250 i.e., £500 millions.

SECOND FORM

		User of Output		
		Steel	*Coal*	*Railroads*
Producer	Steel	·3	·2	·3
of	Coal	·3	·1	·4
Input	Railroads	·2	·4	·1
	Labour	·2	·3	·2
	TOTAL	1	1	1

Here we show that .3 of steel goes to the steel industry, ·2 of steel to the coal industry and ·3 of steel to the railroads industry. So for each rupee worth of steel there is used up 0·3 rupee in steel, 0·3 rupee in coal, 0·2 rupee in railroad and ·2 rupee in paying for labour. Similarly, we can interpret the figures for coal and railroads. Now if the outputs be worth S, C and R in the three industries, and if we want respectively portions s, c and r for use by consumers, we can write for steel

$$S = \cdot 3S + \cdot 2C + \cdot 3R + s$$

Similarly, for the other two we can write

$$C = \cdot 3S + \cdot 1C + \cdot 4R + c$$
$$R = \cdot 2S + \cdot 4C + \cdot 1R + r$$

For this we shall need labour sufficient to earn (at an assumed fixed wage, w)

$$\cdot 2S + \cdot 3C + \cdot 2R$$

which is the total wage-bill. Hence the supply of inland labour should not be less than

$$\frac{1}{w}\,(\cdot 2S + \cdot 3C + \cdot 2R)$$

If we draw up such an input-output table, and calculate the different requirements per unit of various outputs, we can study what the outputs of steel, coal and railroad service should be to ensure supplies indicated above by s, c and r.

The type of figures shown in the above table are called *technical coefficients.*

THIRD FORM

In billions (10^9) *dollars*

Sector of the economy	*Gross out-put*	*Inter-sector flows* 1	2	3	4	5	6	7	8	*Net out-put*
1	17·0	—	—	—	—	0·6	—	0·6	0·7	15·1
2	3·8	0·1	—	1·2	—	—	—	1·3	0·9	0·3
3	12·2	0·7	0·1	0·0	0·3	0·1	0·3	2·1	4·2	4·4
4	8·9	0·4	0·3	0·4	—	0·1	0·3	0·4	2·6	4·4
5	7·0	0·1	—	0·3	—	—	—	0·1	0·8	5·7
6	4·3	1·3	0·3	0·4	1·0	—	—	0·5	—	0·8
7	19·2	0·9	0·1	0·4	1·0	0·5	0·6	—	5·2	10·5
8	60·9	8·2	1·5	3·4	3·1	3·1	0·7	9·1	—	31·8

In this table there are shown eight sectors of the economy of the U.S.A. The figures show the flow in billion dollar values of the gross output of each sector to the eight sectors leaving a net output at the end. The eight sectors are numbered 1, 2, 3, 4, 5, 6, 7 and 8 which indicate them as follows:

1. Agriculture and food industry
2. Mineral industry
3. Metal industry
4. Power industry
5. Textile, leather and rubber industries
6. Railway transport
7. Chemical, wood and paper industries
8. Other industries

The portion of table *minus* the first two columns and the last column is said to constitute the *matrix of inter-sector flows.* In the above table, the matrix has eight rows and eight columns. If we divide the economic activities in an economy into n sectors, the matrix of inter-sector flows will be a square matrix with n rows and n columns.

The whole table, including the gross output and net output columns, is called the *expanded matrix of the balance of production.*

The net output is sometimes called *final output* or *final demand.*

For each sector, we can write an *equation of allocation of output.* Thus, for the third sector we have

$$12{\cdot}2 = [0{\cdot}7 + 0{\cdot}1 + 0{\cdot}0 + 0{\cdot}3 + 0{\cdot}1 + 0{\cdot}3 + 2{\cdot}1 + 4{\cdot}2] + 4{\cdot}4$$

If we indicate the gross output by X_3, net output by x_3 and the other eight flow values by x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37} and x_{38}, we would write the above numerical relationship algebraically as follows:

$$X_3 = \sum_{i=1}^{8} x_{3i} + x_3$$

where $\sum_{i=1}^{8} x_{3i}$ stands for the total of all, x_{31}, x_{32}, . . . and x_{38}

It will be called the equation of allocation of output for the third sector or the *balance equation* for the third sector.

Thus, in the above example, there will be eight balance equations.

However, in general the economy is divided into n sectors, and there are n balance equations for the various sectors of the national economy:

$$X_1 = (x_{11} + x_{12} + \dots + x_{1n}) + x_1 = \sum_{i=1}^{n} x_{1i} + x_1$$

$$X_2 = (x_{21} + x_{22} + \dots + x_{2n}) + x_2 = \sum_{i=1}^{n} x_{2i} + x_2$$

. . . .

. . . .

$$X_n = (x_{n1} + x_{n2} + \dots + x_{nn}) + x_n = \sum_{i=1}^{n} x_{ni} + x_n$$

The algebraic presentation of the n-sector table is given below:

Sectors	*Gross output*	1	2	3		n	*Net output*
Labour	X_0	x_{01}	x_{02}	x_{03}	...	x_{0n}	x_0
1	X_1	x_{11}	x_{12}	x_{13}	...	x_{1n}	x_1
2	X_2	x_{21}	x_{22}	x_{23}	...	x_{2n}	x_2
3	X_3	x_{31}	x_{32}	x_{33}	...	x_{3n}	x_3
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
n	X_n	x_{n1}	x_{n2}	x_{n3}	...	x_{nn}	x_n
Profit		m_1	m_2	m_3	...	m_n	
TOTAL		X_1	X_2	X_3		X_n	

Sometimes to the expanded matrix of the balance of production is added

(*i*) a row at the top showing the distribution of labour force expressed either in man-days or in total wage-bill paid by each sector. Algebraically, this may be indicated as follows:

$$\text{Total labour force, } X_0 = \sum_{i=1}^{n} x_{0i} + x_0$$

where $\sum_{i=1}^{n} x_{0i}$ is to be interpreted as(say) $\sum_{i=1}^{n} x_{3i}$ above and x_0 stands for rest of labour. This is called the *balance equation of labour force.*

(*ii*) a row at the bottom showing the profit for each column sector. This makes the column total (*if all entries are in money values*) equal to gross output. If we use $m_1, m_2 \dots, m_n$ for profit from the third sector column we can write:

$$X_3 = x_{03} + (x_{13} + x_{23} + x_{33} + x_{43} + x_{53} + x_{63} + x_{73} + x_{83}) + m_3$$

or, in general, $X_3 = x_{03} + (x_{13} + x_{23} + \dots + x_{n3}) + m_3$

where x_{03} is the total wage bill, m_3 the profit, and $x_{13}, x_{23}, \dots, x_{n3}$ the amounts spent on supplies taken from the sectors $1, 2, 3, \dots, n$.

The total cost of the output X_3 will be $X_3 - m_3$ or

$$x_{03} + \Sigma x_{i3}$$

However, since we can express X_3 in two ways, we can equate the two and write

$$\sum_{i=1}^{8} x_{3i} + x_3 = X_3 = x_{03} + \sum_{i=1}^{8} x_{i3} + m_3$$

Now, Σx_{3i} contains a term x_{33}: Σx_{i3} also contains x_{33}
This is the portion of X_3 used up in the third sector itself. If we eliminate it from both sides we can write

$$\sum_{i\neq 3} x_{3i} + x_3 = x_{03} + \sum_{i\neq 3} x_{i3} + m_3$$

In general, there will be n such relationships, one for each sector. Therefore, in general, for a sector (say, the jth sector) we may then write

$$\sum_{i\neq 3} x_{ji} + x_j = x_{oj} + \sum_{i\neq j} x_{ij} + m_j$$

These n relationships are called *equations of flow equilibrium.*

Since out of x_3, $\sum_{i\neq 3} x_{i3}$ is paid to other sectors as cost of inputs, only $x_{03} + m_3$ is the value by which $\sum_{i\neq 3} x_{i3}$ is increased. $x_{03} + m_3$ is called the *value added* by sector 3. In general, we can state that for each sector—

Sector output = (*i*) flow to other sectors plus final demand
= (*ii*) flow from other sectors plus value added.

Incidentally, sector-output can also be looked up as being equal to flow from other sectors + value of labour power used up + profit and this can be compared with Marx's division

$$c + v + m$$

where c = value of means of production used up
v = value of labour power used up, and
m = surplus value or profit.

The table opposite is very much like a social accounting table. It is divided into a number of cells; and each cell can be conceived as having an entry, though in practice it is not necessary. The table shows for a sector (say, sector 3) not only how much of its gross output flows to the 11 different sectors but also what flows to exports, stocks and capital formation, government and consumers, i.e., the net output or final demand sector is broken up into a number of important subdivisions.

Similarly, the column for sector 3 shows not only the sector inputs but also inputs in the form of imports, government share (net indirect tax payments), primary inputs for labour, interest, rent and profit.

It must be mentioned that the capital formation column includes all fixed capital and net stock (inventory) changes. It sometimes includes government capital formation (as in the U.K.) and sometimes not (as in the U.S.A.).

The government column includes consumption by government, as also capital formation by government if not excluded from the previous column.

FOURTH FORM

Input from industries	*Output to industries* — *Numbered as for inputs* 1	2	3	⋮	⋮	11	*Total*	*Final demand* — *Exports*	*Capital formation*	*Govt.*	*Consumption*	*Total gross output*
1			2									x_1
2			12									x_2
3	10	3	—			12	180	9	280	56	53	375
4												
5												
6												
7												
8												
9												
10			46									
Not allocated, 11			7									x_n
TOTAL			158				1318					X
Imports			3					e (Reexports)				M
Government			15								DT	
Primary inputs			189				*GNP* at *f.* cost					
Gross output	X_1	X_2	375			X_n	$X = 6494$	E	Sk			

The consumption column includes what goes to the consumers.

On the row side, it may be mentioned that the government row generally includes net indirect taxes (net of subsidy) but sometimes (as in the U.S.A.) it includes all taxes and even some miscellaneous payments to government so far as the industrial sectors are concerned. The government-row entry (DT) under consumption shows direct taxes paid net of transfer payments.

The row for primary inputs indicates what individuals get as labourers, organisers, capitalists, entrepreneurs and windfall receivers.

Incidentally, the import-row entry (e) under the export column refers

to reexports. Also, the primary-input-row entry in the total column shown just after the industry-sectors refers to gross national product at factor cost.

The whole table can be summarised as a two-fold table as indicated below:

			TOTAL
	Inter-industry transactions not entering national income estimates 1,318	*Final demand* 5,176	6,494
	National product and import 5,176	*Reexports and direct consumption of imports* 438	5,614
TOTAL	6,494	5,614	= 12,108

The top left-hand cell is really the inter-industry or input-output matrix. But, as is evident from the four forms, there is a variety of presentations used in writings on input-output analysis.

It may be added that these relate to current inputs as against capital inputs. Thus there are two input-output tables:

(*i*) Input-output table for current transactions

(*ii*) Input-output table for capital transactions: the word *capital* refers to durable inputs used in industries.

General Concept of Input-Output Table with Comments from Economics Viewpoint

Instead of the usual factors of production we conceive of more categories of goods and services used and call each an input. These inputs come out as produce (or output) by man. Even man can be looked upon as a category, which enters different production channels as input and for which he appears as consumer of net output. A two-way n-sector input-output table is conceived as follows:

Input sectors	*Outputs to*					
	A_1	A_2	A_3		A_{n-1}	A_n (Consumers)
A_1	x_{11}	x_{12}	x_{13}	...	$x_{1(n-1)}$	x_{1n}
A_2	x_{21}	x_{22}	x_{23}	...	$x_{2(n-1)}$	x_{2n}
A_3	x_{31}	x_{32}	x_{33}	...	$x_{3(n-1)}$	x_{3n}
...						
...						
...						
A_n (Labour)	x_{n1}	x_{n2}	x_{n3}	...	$x_{n(n-1)}$	x_{nn}

Each of the A_1, A_2, A_3,...categories is called a sector. Each sector stands for a distinct homogeneous (and if necessary, homogeneous class of) economic activity. Thus one sector may refer to 'agriculture'; another may refer to 'power' including gas and oil as well as electricity.

There are in general $n \times n$ entries in the table. But many of them are likely to be blank. Thus the output of mining may go to a few output-sectors while the output of power may go to almost all.

Each row shows the distribution of output of each sector.

Each column shows the inputs used up in the production activity of its particular sector. The inputs may be all shown either in physical quantity units or in value terms.

The A_n-column is a residual column, a net-output column receiving what is not used up in industries. These go to produce men whose distribution as labour force in man-days (or money-remuneration) is shown in the A_n-row.

If we conceive of such a tabular presentation, then the following remarks about it deserve our attention:

1. The entries relate not to a point but to a period (say, a year) of time.
2. The unit of the n horizontal entries against any (say, A_3) row is the same; it is the unit in which A_3 is measured. It is a physical unit. So the unit *may* vary from row to row but is the same within each row.
3. The units of the n vertical entries in a column against (say, A_3) are different as these entries refer to different goods going into the production activity of the sector called A_3. The vertical entries are costs for the total of entries in the respective rows.
4. The entries against each sector are empirical and not theoretical.
5. The number of sectors are not as many as the goods produced in an economy. So each sector stands for a basket of goods. These goods should in general be homogeneous, e.g., A_3 may be for coal of all types. But in the present-day system of production, data become available not so much in terms of goods as in terms of production units. So the horizontal entries against the sector A_3 may stand for goods produced by the coal mines. These goods need not be coal only. Each sector includes by-products as also other main products. Vertically, against A_3, we shall learn of the quantities of the various goods that get used up in the production of what is shown in the A_3-row.

Where separate data are available, certain products (output) of A_3 may be included for distribution purposes in other (say A_4 and A_5) rows. If coal mines keep complete separate accounts for the production of the byproducts or joint products we may distribute the vertical entries also into the appropriate columns. In practice, this is seldom likely to be so. So, while there is a feasibility of distributing outputs into appropriate rows, the cost items do not admit of such adjustments.

Such horizontal (but not vertical) adjustment is likely to distort the cost picture of A_3. The vertical entries of A_3 will become cost entries for what is shown as distributed against A_3 in the row. This will increase the cost of production of A_3, and reduce the cost-items of A_4 and A_5 outputs as shown in the input-output table.

6. Inputs are determined by production requirements of a unit: they are technologically determined: they are bound up with the techniques of production actually adopted. If production continues on the same lines and in the same pattern the pattern of inputs is not likely to change. But does this imply production at maximum (or optimal) efficiency? Does it mean that there is no further scope for improving efficiency of production by economising on inputs or by increasing output? The answer to both questions is in the negative. The input-output table as such does not admit of such considerations.

If we could draw up input-output tables for different times and make a comparative study, we may be able to analyse or understand the scope for more economy, efficiency and substitutions.

If we divide each of the input-items shown against A_3 by the total output of A_3 we get the inputs required per unit of A_3-output. These are called *technical coefficients* of inputs for A_3-output. But will it be the (*i*) correct, (*ii*) average, or (*iii*) marginal requirement? It may not be the correct requirement if the input is under-utilised in A_3. It is not likely to be the marginal requirement as it is found from total output. It is therefore of the nature of average requirement, subject to the correction-remarks above. For the future, it can be taken to be the average requirement if we decide to think that, in the future also, this input will be utilised in the manner and to the extent it was utilised during the period to which the input-output table relates.

The position is therefore likely to be misleading in regard to durable capital requirements.

In any case, the use of the technical coefficients as derived from the input-output table assumes

(*a*) full use of inputs
(*b*) linear constant returns to scale

without possibility of input-substitutions.

Hence changing prices are not likely to affect the relative proportion of inputs.

7. The use of inputs shown in the A_3-column for labour is particularly subject to effects of

(*a*) income and income-distribution among the people,
(*b*) inter-personal consumption-demonstration effect,
(*c*) consumers' preferences,
(*d*) prices

8. It will be difficult to make the entries where inputs are in the form

of services such as banking, insurance, transport and ministerial services. These can be either left out or evaluated in terms of value or price of service rendered to different sectors.

9. A_n, i.e., man himself, decides what should be termed an optimum use of $A_1, A_2, \ldots, A_{n-1}$, (nay) even of A_n. If he decides to have a 42-hour week for himself, then what he does during the remaining hours is not reflected anywhere. The input-output table is not a schedule showing the 24-hour activity of the sectors: nor does it give any idea of the unused capacity, i.e., of balances of assets and stocks; nor is there a place in it for stocks, shares and bonds.

10. It is now usual to draw up the input-output table in money terms. This no doubt makes the technical coefficients subject to changing prices in the economy.

11. The entries in the table do not distinguish between constant (durable) and variable inputs. Of course, where equilibrium exists, the flow of durable inputs should be equal to what are used up by way of depreciation and wear and tear in the sectors. In that sense even, the durable input entries should be akin to variable input entries.

Input-Output Analysis

In conclusion, let it be clear that input-output analysis is conceived not so much as a means to the establishment of an equilibrium except perhaps apparently, as (*i*) to secure an imbalance (where output exceeds input) to provide for beyond-the-control increases in population, labour force, balance of trade deficits, etc. or (*ii*) to ensure an increasing positive imbalance (excess of input over output) for each individual or individual-household living in a society (community or country). Such a state, even when achieved, cannot be permanent, although it may continue to exist for a considerable length of time (*i*) by drawing upon internal (inland) resources; or (*ii*) by using resources robbed, mobilised or secured from external (alien) regions. It cannot be permanent because fundamental sources (of resources) in their present form must get exhausted one day, so as to lead to the start of a reaction and of the strengthening of forces that would reverse the imbalance.

CHAPTER 14

Welfare Economics

Meaning of Welfare

The economist knows of two states, namely, the welfare state and the illfare state. When a man has a want he suffers pain and when that want is satisfied he experiences pleasure. These feelings of pain and pleasure are associated with the words illfare and welfare. Some wants always remain unsatisfied, giving rise to illfare, and some wants are almost always satisfied or are in the process of being satisfied, thus giving rise to welfare. Welfare can, then, be understood as consisting in that state of the mind which results from want of satisfaction. In whatever other sense the word welfare may be used, the fact remains that welfare is the result of the removal of wants. And it can be made to follow from this fact that to increase the welfare of a man his wants (some or all, if at all that be possible) must be satisfied.

Comparing two periods of time we can say that a given man has greater welfare in that period in which a larger number of his wants have been or are being satisfied. Conversely, we can say that the smaller the number of wants (of given intensities) that remain unsatisfied the greater is the welfare.

Can Welfare be Immoral?

An answer to this question necessitates knowledge of the meaning of the

word *immoral*. For most of us immorality consists in doing good to oneself at the cost of others. This meaning of immorality is, for most practical purposes, correct. But there is a way of defining this word without bringing in directly the effect of one's action on other people. We can say, for instance, that an immoral act is that act of a man which, in the totality of its effect, is against his own interest. The totality of effects naturally spread themselves over a long period of time. And in order to make this definition foolproof we must let time extend itself beyond the grave. In other words, the effects of an immoral act can be experienced by one even in one's life hereafter. In simple words, while an immoral act is that which secures one's benefit at the expense of others, it is also that which, in the ultimate analysis, is against one's own interest.

We are now ready to answer the question posed above. When a man satisfies his wants (for, that is the way of securing welfare) at the cost of other people his welfare is immoral. Alternatively, his welfare is to be called immoral when the manner of getting it decreases it in the long run. To remove the dichotomy of welfare into moral and immoral we must, therefore, take a long period of time into account. But if we stick to the commonly accepted meaning of immorality, the distinction between moral and immoral welfare will have to be retained. And then moral welfare would be that which does not harm the interest of other people or, since we live a socially integrated life, it is perhaps that which actually increases the welfare of others. This takes us to the examination of the concept of social welfare.

Welfare Economics and Social Welfare

In welfare economics we do not study the economics of welfare for the reason that that is the only economics that we know of—economics is naturally and always economics of welfare. What we really study under the title welfare economics is the economics of social welfare. The first point to clarify, therefore, is the difference between individual and social welfare. Without doubt, social welfare is the welfare of the whole society and, therefore, it would appear there is a quantitative difference between social welfare and individual welfare. All those who attempt to find criteria for maximisation of social welfare tacitly assume that, in some sense, it is the aggregate of the welfare of all the individuals composing a society. Take, for example, the Paretian basic proposition that, when all the individuals are better off (or some are better off and nobody is worse off) than before, social welfare must be considered as having increased. The entire structure of welfare economics since the time of Pareto rests on this fundamental proposition. Now, the only logically permissible statement should be that when everybody is better off every-

body is better off. To jump to the statement that, therefore, social welfare is greater involves an unexamined concept of social welfare, perhaps that it is the sum total of the welfares of all the individuals. For, if social welfare is the summation of individual welfares, it would certainly follow that any addition made to the welfare of any individual would inflate social welfare.

A scientist must define every term that he uses. The word social welfare must first be defined before we proceed to determine the criteria for its maximisation.

The Concept of Social Welfare

Since the adjective *social* is derived from the noun society we must know what a society is. Without beating about the bush, let us say that when people help one another to satisfy their wants they are said to form a society (in respect of the satisfaction of those wants). The wants that they co-operate to satisfy may be called socialised wants. Since all the wants are never socialised, people can never be taken to form a society irrespective of their wants. Consequently, social welfare should include only that welfare which is derived from the satisfaction of socialised wants. This is what logically follows from the above concept of society. If, therefore, some people are better off because they get more satisfaction than before from the removal (satisfaction) of wants that are not socialised (wants which they satisfy by their individual efforts), social welfare cannot be said to have increased. For practical purposes, it is assumed (without the economists being conscious of the fact that they are making an assumption) that all the wants of the people are socialised. If and when all the wants are socialised, every addition to the welfare of an individual would constitute a component of social welfare. Yet, it remains to be seen what precisely is social welfare and where its residence is located. Being welfare, social welfare cannot exist anywhere but in the human mind and since the human mind is possessed by a human being it follows that social welfare is experienced by an *individual* (and therefore by all individuals). We shall have occasion to revert to this point in the following section.

From Pareto to Modern Economists

Pareto's statement in regard to social welfare is comparatively innocent. He says that social welfare is increased when some at least are better off without anybody being worse off. The optimum condition that is derived from this is that social welfare is maximised when it is not possible any further to make anybody better off without making others worse

off. *Assuming* (though, as said above, we cannot logically assume that) *that every addition to the welfare of individuals makes an addition also to social welfare,* whatever its precise nature may be, Pareto's proposition is one to which perhaps no objection can be taken. On this pedestal rests the entire structure of welfare economics as it has developed since the time of Pareto.

To one important point attention may be drawn at once. The fact of being better off is a function of two variables. First, it is a function of an individual's own state, a function of what he has. Second, it is a function of what others have. In other words, it is a function of the absolute and relative status of an individual. Thus, when we make a statement in regard to welfare couched in terms of such terms as *better off* or *worse off,* no separate consideration of *distribution* is needed. The state of being better off takes account of the relative position of the individual and so of distribution.

But when we speak in terms of a *bundle of goods and services* that an individual prefers, separate mention has to be made of the distribution pattern in the society. Thereby some complications creep into our analysis of the problem, making us feel that it involves value judgment.

PLACE OF VALUE JUDGMENT

When the value of a thing is not derived by a process of logical reasoning from that of another thing it is said to constitute a case of value judgment. Thus, it is believed that it is not possible to say on the basis of any logical reasoning whether one pattern of income distribution or wealth distribution is more desirable than another. Somebody has to *give* us a distribution as one that is best and there can be argument on that question. If one distribution is, therefore, preferred to another, for purposes of determining social welfare, it rests on a taken-for-granted judgment in regard to certain values. And when we admit the involvement of value judgment the magnitude of social welfare becomes a function of a variable that has no economic significance (making a distinction between economic and non-economic variables).

Thus, the inclusion of value judgment robs welfare economics of logical and scientific precision. When, however, we couch our propositions in terms of more inclusive words such as *better off* and *worse off* we bypass the difficulty that we otherwise inevitably encounter.

If, however, social welfare is taken to depend on distribution (thus dragging value judgment into the picture) it is very pertinent to ask: *Who is the recipient of that welfare?* If a certain pattern of distribution is accepted as the best, he who accepts it as such will, in the event of its being achieved, experiences the welfare in its maximised intensity. But to make welfare an objectively maximised feeling it would be necessary to suppose that a certain pattern of distribution is accepted by all as

the most desirable one. Now, if there are two components of welfare, one, the welfare experienced by individuals and derived from the absolute quantities of goods and services enjoyed by them and, two, the pattern of distribution, (social) welfare becomes a queer mixture of the independent judgment of individuals and the judgment of an exogenous agency that is forced on them for acceptance. If the value judgment is the product of the working of the mind of a superman, social welfare is unequivocally experienced by the superman only, unless there are others who fall in line with him. But going back to Pareto's fundamental proposition, we know that if it has to be logically correct, every addition to the welfare of individuals must cause also an increase of social welfare (though not necessarily to the same extent). The effect of the pattern of distribtution on the welfare of individuals thus becomes a causative factor in the make-up of social welfare. Each individual will make his own judgment about the value of a given distribution-pattern and accordingly allow it to influence his welfare. If there is, however, one particular pattern of distribution that is taken to be desirable it is the one that is forced upon the judgment of all. If it is not so forced upon them the social welfare that results from it has nothing to do with the welfare that individuals experience.

The Residence of Social Welfare

All that has been said above should make it clear that social welfare is not a uniquely determined and uniquely experienced feeling. Social welfare, by virtue of the fact that it is welfare, must have its residence in the mind of an individual (and, therefore, in the mind of every individual). There, in his mind, live peacefully two species of welfare, individual or private and social. And for a distinction, separating the one from the other, we have to have a dichotomy of wants as mentioned earlier. There are wants that individuals satisfy individually (of course operating in an environment that is provided by natural and man-made forces) without any idea of helping others or taking help from them. There are also wants that they help one another to satisfy. The welfare derived from the satisfaction of the former category of wants may be (should be) called individual welfare while that derived from the satisfaction of the latter category may be (should be) called social welfare.

Social welfare then, and so its measure, is not a unique entity; it differs from man to man. While one may feel that social welfare is greater this year than what it was last year, others might feel it is less. This seemingly strange statement about social welfare is, however, one that follows logically from the basic propositions in regard to welfare. In the case of other concepts of social welfare, the superman (or one who is

so regarded) becomes a dictator, forcing his judgment of welfare on the people. These considerations point to one indisputable fact, namely that it is futile to search for some objectively determined criteria of social welfare or even criteria that do not differ from man to man.

When Individuals Move in Different Directions: the Compensation Principle

Underlying Pareto's basic proposition is the assumption that every addition to the welfare of individuals causes an increase of social welfare. From this it should follow that every decrease of the welfare of individuals decreases social welfare also (though not necessarily in the same proportion). For the calculation of social welfare, or to determine the direction in which it has changed, one has to take account of both increases and decreases of the welfare of individuals. When one has thus to take positive and negative quantities into one's reckoning some way has to be found of striking a balance and determining the net result. If by some process the increases and decreases of welfare could be resolved into only an increase or only a decrease of welfare, Pareto's proposition could be used to make a statement about social welfare.

The above considerations take us to the famous compensation principle of Kaldor and Hicks. Suppose that, after a reorganisation of the economy of a country, some individuals find themselves better off and others worse off (the modern economists express the same idea by saying sometimes that some individuals find themselves on higher indifference curves and others on lower indifference curves). In such a situation, no definite statement can be made about welfare. For, as Pareto said, to be able to say whether social welfare has increased or decreased either all the individuals should be better off (or some better off without others being worse off) or all should be worse off (or some worse off without others being better off). Hence, a device is needed to convert the situation into one in which the welfare of individuals moves in the same direction. N. Kaldor, therefore, introduced the compensation principle by saying that in such a case the gainers should be able adequately to compensate the losers if social welfare has to be treated as having increased. There has to be paper calculation only; no actual transfer of income is needed. We have to find out what amount of tax is needed to take the gainers back to their original position and what amount of bounty is necessary to push up the losers to their original position. If the former amount (tax) is greater than the latter (bounty), social welfare can be considered as having increased.

This principle, to which J. R. Hicks gave his support, has certainly something to commend itself. There is an attempt in this principle to

convert the economic position of individuals in such a way as to make it possible for us to apply the Pareto criterion. However, it has been objected that, according to this principle, one can only say whether there is potentiality of greater welfare in the reorganised economic system. For, unless taxes are actually levied and bounties paid, social welfare remains what it was (making it impossible for us to say whether it has increased or decreased).

There is much in the above objection but it is attempted to be met by Hicks by maintaining that, from the purely economic point of view (perhaps also from the macroeconomic point of view), the reorganised state can be considered as one in which there is increased welfare. It remains, then, only for the state to effect whatever marginal adjustments are felt necessary to make the increased welfare position effective in making everybody feel better off. Since social welfare is a macro concept there appears to be some substance in Hicks's contention. And his contention would gather further strength if the view of social welfare expressed by us earlier is taken.

However, at this stage, we may leave this point to the judgment of the reader. The whole trouble arises from the fact that there never is a real macroeconomy in this world. If it could exist, Hicks's contention would perhaps become quite valid.

Dual Aspect of Distribution

Compensation from the gainers to the losers causes a redistribution of income. This redistribution effects a change in the position of individuals as far as their incomes go and, by so doing, alters the welfare of the people derived from their absolute (not relative) economic status. This is one aspect of distribution. But the pattern of distribution of income, either before or after the losers are compensated by the gainers, has its impact on the relative economic position of individuals. The welfare of the people is a dual function of distribution. In the first place, it has an effect on the absolute position of individuals and, in the second, on their relative position. We might, for the sake of simplicity (to avoid complications caused by the fact of value judgment), ignore the second aspect of distribution. But it is not possible to ignore the first aspect while applying the compensation principle.

Concentrating attention on this first aspect of distribution, let us note that some critics of the compensation principle as stated by Kaldor and Hicks object to their considering the possibility of redistribution of income after reorganisation of the economy of a country and ignoring such a possibility before reorganisation. Here the name of Scitovsky must be mentioned as a strong critic of the Kaldor-Hicks compensation

principle. He rightly pointed out that to judge which state is better from the point of view of social welfare we must consider the possibility of redistribution in both the states. Further refinements called for by considerations such as this one came from Samuelson. Since then, the history of the development of welfare economics is a long and complicated array of theorems, most of which perhaps add little to the substance of old welfare economics.

A Composite Social Welfare Function

We shall now choose to be a little more theoretical. And in becoming more theoretical we shall, in a way, come close to the concept of social welfare that we placed before the reader earlier in this chapter. Bergson formulated the social welfare function in 1938. Since then, the idea has been grasped and made use of by other economists. Arrow extended its scope in an attempt to explore the possibility of such a function being meaningfully constructed for a society of free individuals.

For purposes of theory, one can imagine social welfare as a state of the mind of some particular individual (say, the superman) or of the minds of all individuals, if they think and feel alike, or of the mind of an average individual. In the first case, logically, social welfare is the welfare of the superman; in the second case, it is the welfare of everybody; while, in the third case, it might not be the welfare of any particular individual since the mathematical average man may not exist in flesh and blood.

The superman may think that social welfare is maximum when a particular use of resources is made and the result of it becomes available to individuals according to a certain scheme of distribution. In his opinion, social welfare is then maximised—it is maximised because he *feels* the entire economic position to be most satisfying. But then it is *his* welfare and nobody else's, though it is labelled as social welfare. It is his estimate or his feeling of social welfare, the location of which is perhaps not known to him. But what are the characteristics of a superman? If he is a superman in the sense that his sensitive heart records all the feelings of joy and sorrow of all the people, social welfare as recorded by him becomes a sumtotal of the welfares of all the individuals in a society. This would certainly be a dubious concept of welfare. For, social welfare would then be greater (other things being the same) the larger the population.

If, on the other hand, the superman does not simply record the feelings of others and allows them to accumulate in his heart, but so accommodates them as to make social welfare as experienced by him the welfare of individuals combined into a true macroeconomy, the position

macroeconomy is never to be found in this world, the superman's feeling of welfare does not project itself into the hearts of other people.

Nor can social welfare be conceived of as the welfare (the state of the mind or of the heart) of all the individuals, for the reason that all cannot think and feel alike. In fact, it would be difficult to find two persons who are alike in respect of all those factors on which welfare depends. A similar difficulty arises when one thinks of social welfare as the welfare of the average man. The average man is a mathematical concept that does not have its counterpart in the world of flesh and blood.

Any attempt to construct a social welfare function would have to encounter the difficulties mentioned above. Once we assume, for purposes of analysis, that social welfare is the welfare experienced by the superman, it becomes possible to make it a function of all those variables that affect the welfare of the superman. At any rate, such a mathematical device is possible in theory. We can similarly substitute the average man for the superman. But, whatever we do, welfare remains the welfare of a particular man (or of a particular type).

A Pragmatic View of Social Welfare

Is there anything like social welfare then? Suppose a society consists of two persons. If both of them are found to be more satisfied today than they were yesterday a neutral person would express this phenomenon in a straightforward way by saying that both the persons are better off today as compared to what they were yesterday. If, calling the two persons together a social organisation or simply a society, one were to say that the society is better off it would be treated as a short form of the statement that both the persons are more satisfied today. In the same way, when we say that social welfare increases when we spend more money on health services, all that we intend to say is that a large number of people get the benefit of health services. Similarly, when we educate poor people and provide them with the bare necessaries of life, taxing the rich for the purpose, we might maintain that social welfare has increased. As a matter of fact, all that happens in such a case is that a large number of people get the benefit of education and have their primary wants satisfied at the expense of a comparatively small number of people. To draw the conclusion from this that social welfare has increased is unscientific. But that is the pragmatic view of social welfare. As a matter of fact, so many changes take place simultaneously in an economy—those who benefit from government activities do not benefit to the same extent and those who suffer do not suffer equally—that it is impossible to say whether social welfare (even from a pragmatic point of view) has increased or decreased.

The difficulties alluded to above get enormously magnified when one takes long-period point of view. What is good from the short-period point of view is not necessarily so in the long run. Further, wise men claiming to judge the welfare of a society are bound to hold different views on the matter and set different goals before the government for the achievement of maximum welfare. The task is beyond the mental and psychological capacity of any individual to accomplish. It has, therefore, been well said that we should stop spending more time and energy on finding criteria of maximum social welfare.

Economic System and Welfare

One of the determinants of welfare is the economic system. The welfare of the society (whatever that might imply for one) is differently determined under capitalism, socialism and communism. In the case of capitalism it will vary according to the relative importance of pure competition and monopoly. Further, the effects of competition on welfare will depend on the technique of production that determines the extent of external and internal economies.

It is possible to make one very general statement in regard to the economic system of a country. The gap between capitalism and socialism (or communism) is greatly narrowed down when competition becomes pure and perfect. The disadvantages of competition spring not from the fact that it is competition but from the fact that it is seldom, perhaps never, real competition. All the criticisms of competition, if carefully examined, would be found to be based on the fact of imperfection or impurity of competition. The idea behind competition is to prevent one party prospering at the cost of another. If this could be done through the perfection of competition, we would, to a certain extent, have the ideals of socialism and communism realised. We shall leave this point here with the observation repeated that competition would become desirable if only it could be made pure and perfect. Capitalism would also become desirable if it could create conditions in which competition could function without any economic, social or natural restraints.

In traditional economics, one finds certain conditions laid down for maximum social welfare or what is sometimes called the Pareto optimal. These conditions relate to the production and consumption of goods (services included) on which welfare is taken to depend. One might say that welfare is maximum when production and consumption are most efficient, i.e., when it is not possible any further to shift any input from one plant or one industry to another without decreasing the output anywhere, and similarly when it is not possible to shift goods from one consumer to another without making the former worse off.

This latter requirement, stated as it generally is, is logically unsatisfactory. If we reallocate consumption goods among consumers so that somebody becomes worse off but at the same time somebody else becomes better off, welfare can be taken to have increased according to the traditional way of thinking. It is, therefore, not only at the optimum level that a transfer of consumption goods from one consumer to another decreases the utility of the former. However, cover can be taken under the plea that we are concerned with the conditions of the Pareto-optimal. It is difficult, however, to imagine any case in which reallocation of goods for consumption purposes does not make somebody worse off. If the traditional view of social welfare is taken, a shift of consumption-goods from one consumer to another increases welfare if it deprives the former of less utility than it gives to the latter. Here, it is believed, inter-personal comparison of utility is involved. We do not think so, however. But our objection to the very concept of social welfare stands in the way of our accepting the above statement. Indeed, the condition of the Pareto-optimal, if it is worded as above (as far as consumption goes), tacitly assumes that social welfare is in some sense an aggregate of individual welfares or that it makes no difference to social welfare to whom the utility of a given intensity accrues.

The condition of the Pareto-optimal, as far as production goes, is more innocent. Other things being given, social welfare is greater (certainly not less) when production is greater. It is hard to keep other things the same, but given that assumption, the condition seems to be acceptable.

Now, the condition of optimality relating to consumption turns out to be a question of distribution. Having produced goods and services, they have to be optimally distributed. But production and distribution are not quite independent of each other. The technique of production and the system of remuneration determine also the distribution pattern which, however, can be modified by governmental activities in the sphere of public finance. Normal finance and functional finance both alter the prevailing pattern of distribution.

The above is the first aspect of the distribution pattern. For, distribution also has another aspect: it determines the relative as well as the absolute income and, therefore, the status of an individual. The second aspect of relative status offers more difficulty than the first. Each individual feels better-off or worse-off according as his relative income is greater or less. The extent to which he feels better or worse depends on his psychology and, therefore, varies from man to man. There is, in this effect of distribution, no value judgment involved. But value judgment comes into the picture when an independent observer looks at the economic condition of the people and makes his estimate of social welfare. In the discussion of welfare criteria, it is thought wise, therefore, to leave the question of distribution, in as much as it involves value judg-

ment, apart and confine attention only to the rest of the variables on which welfare depends.

COMPETITION AND WELFARE

It was observed earlier that, when competition is pure and perfect, the resulting economic order corresponds to that under socialism and communism in some of their essential features. Competition, it was pointed out, prevents the exploitation of one party by another. If, by preventing exploitation, social welfare could be optimised, competition may well guarantee those adjustments on the production and consumption sides that lead to Pareto-optimality. It is difficult to visualise an economic order in which competition is not only pure and perfect but is all-pervading, spreading its tentacles wide and deep, all over the entire system. But, in theory, the possibility of such competition does exist and, therefore, also the guarantee that it would secure maximum social welfare. And, by contrast, monopoly, if it could penetrate into every nook and corner of the economic system, would yield a situation of minimum social welfare.

But competition, in the form and to the extent to which it regulates economic activities in this world, does not and cannot always maximise welfare. Where consumption indifference curves are concave to the origin and production indifference curves are convex we do not get a Pareto-optimal even when the rates of substitution are equal. Mathematically speaking, competition by itself ensures the satisfaction of the first-order condition, but for maximisation we need the fulfilment of the second-order condition. If that is not satisfied, competition would yield minimum satisfaction. (For a maximum, the first differential has to be zero and the second differential negative.)

Further, competition fails to secure optimum welfare when there are external economies or diseconomies. The price of a commodity may, under competition, equal private marginal cost but not social marginal cost when there are external effects, as shown by Pigou. In certain industries, production would then be less, and in others greater than what is consistent with the conditions for maximum social welfare. As Pigou argued, in such cases the state has to come into the picture to boost up production in certain industries with the help of bounties and check it in others by means of taxation. The fact is that, in certain circumstances, there is a conflict of interest as between one industry or producer and another. What goes to increase the welfare of one producer damages the welfare of others. This is due to what is technically called external diseconomies. Here too one could say that all such cases are covered by the absence of pure and perfect competition in all the sections and cross-sections of the economy.

Pareto-Optimal Is only Pareto-Optimal

Sensible statements can with confidence be made about social welfare only when individuals and society merge themselves into one. So long as individuals retain their individuality and economic activities are determined, in the first instance, by the maximising behaviour of individuals, social welfare will remain a difficult concept for the economist to use. Social welfare, even when we can manage to show it grown to its full stature, will simply subserve, as it were, the interest of individual welfares. It is for this reason that all attempts to evolve criteria for maximum social welfare have proved abortive. It is only in the case of Robinson Crusoe that sensible statements can be made in regard to social welfare. For, in his case, there is a perfect merging of the individual into society. This point can be illustrated by taking the case of barter.

When *A* has one commodity and *B* another in such quantities that the marginal utility of *A*'s commodity is less than the initial utility to him of *B*'s commodity and, likewise, the marginal utility to *B* of his commodity is less than the initial utility to him of *A*'s commodity, barter of *A*'s commodity (call it *X*) with *B*'s commodity (call it *Y*) gives them consumer's surplus, or, we might say, places both of them on higher indifference curves. There is then a certain rate of exchange, yielding certain quantities to be bartered, that is most acceptable to each of them. Any movement from such an equilibrium position would decrease the welfare of each or, in other words, take them to lower indifference curves. Such a position can then be called the position of the Pareto-optimal. Social welfare is in such a case maximum in the sense, and only in the sense, that any deviation from this position would make at least one of them worse off.

Yet, in such a case of barter, there can be other points at which one of them would be better off, and may be much better off, than before with the other slightly worse off. Whether this is worse than the previous position of equilibrium in which both *A* and *B* are *simultaneously* at their optimum (relative optimum) cannot be said unless we have a satisfactory concept of social welfare. And we have no such concept except (in our view) the one we have mentioned earlier in this chapter. Hence, the Pareto-optimal is only Pareto-optimal.

The confused state in which we find welfare economics today, with all its complications and propositions which are challenged no sooner than they are made, with a bewildering mass of material to which reference has been made by E. J. Mishan, can be attributed to the want of a clear concept of social welfare.

CHAPTER 15

Competition: Pure-Impure and Perfect-Imperfect

Necessary Conditions for Competition

We shall proceed step by step. Without first taking up the question whether competition should necessarily mean pure competition, let us see what conditions must be satisfied if there is to be competition at all. In the first place, there must be at least two parties for the reason that one cannot compete with oneself. In the second place, there should be something to compete for: there should be something which each wants to have. And, as a sub-condition, we may say that the thing for which they compete is not divisible, so that if one secures it the others have to go without it. For competition to exist these necessary conditions must be fulfilled. As we proceed, it will become clear why, for purposes of logically correct definition, the thing that the competitors compete for must be indivisible.

On the face of it, the above language is loose. For, when we say that there should be at least two parties we do not specify within what area. But no specific mention of area is necessary for the reason that when we say that they should be in a position to compete for a given thing it becomes obvious that their location does not need to be bound spatially. So long as they are *able* to compete they give rise to competition, no matter where they are and what distance separates them.

Let the thing for which they compete be called the bone of contention or, for the sake of brevity, simply the bone. In order that the second condition mentioned above may be satisfied it is necessary that there

be no social, natural, economic or legal obstacles to each party's trying to grab the bone of contention. The only obstacle in the way that is permitted by the fact of competition is the effort of rival competitors to prevent one another from laying their hands on the bone.

Divisible Bone of Contention

Logically, there should be only one bone, one thing for which the competitors compete, for each act of competition. In other words, the bone, the object of competition, should be indivisible. Let us, for the sake of simplicity of exposition, assume that there are only two parties. There is then only one thing which each of the two is trying to grab. Now, after fulfilling the necessary conditions of competition we need to have one more condition fulfilled to complete the list of *sufficient* conditions. Not only should the two competitors try to grab the bone, they should also be *able* to grab it. Mathematically, it is impossible for each of the two competitors to get the bone simultaneously. The result then is that, if each has to be able to get the bone, neither of them actually succeeds in getting it.

In the case of competition, therefore, the parties are so matched that none is able to get what he is competing for. Each party is, thus, able to prevent the other parties from securing the object aimed at. This is, logically, the objective manifestation of competition.

Now, to let the case of competition correspond closely to what is witnessed in this world, let us allow the object of competition to be divisible. Since we are planting our feet on the earth, let us take the case of sellers of a commodity. Though they sell a commodity, their object is to capture buyers who constitute, in the language we have used above, the bone of contention. Buyers are obviously divisible and, therefore, if there are two sellers, each of whom has a hold on the buyers, instead of neither of them being able to capture customers, each will have half the number of them. It is assumed that the attitude of buyers is neutral towards the sellers and that the sellers are, by the definition of competition, equally efficient in capturing customers.

Hence, if there are two sellers, competition should result in each seller having half the number of buyers and, if there are a large number of them, the buyers should be equally divided between them.

Stability with Divisible Bone of Contention (Customers)

Will the sellers have equal numbers of buyers from the beginning? And

will they continue to share the buyers equally? If the theoretical conditions of competition are fully satisfied, one should expect equal division of customers from the start till the end. It would be like a tug-of-war in which the two sides are in every respect and at all moments equally strong. But in the world we are studying, competition does not get a chance to attain maturity at once. A situation *grows* into that of competition. And, so long as it is struggling to attain its full stature, fluctuations in the number of buyers that each has would continue. For, each seller would try out various ways of putting his strength to effective use. The chances, therefore, are that the sales of the sellers, together with the prices charged, would continue to fluctuate. But the *norm* would be equal division of buyers—the variations would always be in the direction of equality of custom.

Whether, therefore, there are two or more sellers, even when conditions of competition are satisfied to the extent to which they can be in this world, variations in the division of buyers would be witnessed with a continually disturbed tendency towards equality. This is the reason why, in the solutions of duopoly, we get either a fluctuating price (Edgeworth) or a stable price that is arrived at only under fortuitous circumstances. Extending the case of duopoly to one in which there are a large number of sellers (traditionally called the case of competition), we would get, if we adopt Edgeworth's solution, a continuously and quickly fluctuating price, so quickly and so continuously that it could be regarded as a stable price. Stability is, therefore, hard to imagine when the bone of contention—custom—is divisible. It will be noticed that sellers and buyers have been the operationally significant agents. The fact that they sell a commodity (or a bundle of commodities) has been referred to only in passing simply because a seller is a seller only because he sells something.

Do Competitors Sell the Same Commodity?

We have explained the meaning of the word competition in terms of sellers and buyers with only a passing reference to the word commodity. Since, however, they do sell a commodity (or a bundle of commodities) it appears pertinent to ask what precisely the word commodity means. When are two sellers, for instance, said to be selling the same commodity? This question of the meaning of commodity was discussed at length some years back. For a practical businessman it offers no difficulty. He is not concerned with the scientific meaning of any word. But for an economist it becomes necessary to know what precisely is the connotation of a word that he uses.

Competition, for example, is naturally understood to imply competi-

tion among the sellers of the same commodity. Is a fountainpen of one make the same commodity as a fountainpen of another make? Questions of this type can be multiplied in order to make the concept of commodity made quite clear. Since a commodity is a commodity because it satisfies some want it was felt that two things should be treated as the same commodity if and when they satisfy the same want. Or, it could be alternatively said, with greater logical precision, that two things constitute the same commodity when they give the same utility. But then, what would we say if they gave the same utility to *A* but different utilities to *B*? Naturally, in such a case they are the same commodity for *A* only. It could also be said, though perhaps without much benefit, that to the extent to which they give the same utility they are the same commodity. If one commodity gives 100 units of utility and the other 70 units only, the two commodities can be said to be the same commodity to the extent of 70 per cent. This logically permissible view can be of use to us.

If, then, two things, regardless of their appearance, give the same utility, they must be regarded as one and the same commodity. One point still remains to be clarified. What are we to understand by the word *thing*? Should the word thing have reference only to a material, concrete object?

When a buyer pays the price of the commodity he buys, he gets in return utility not only from the material thing bought but also from the promptness and politeness with which it is sold. Likewise, he gets some added satisfaction when he has to spend less time in approaching the shop. A buyer, therefore, can be said to buy all those things from which he gets utility though, from the business point of view, he appears to buy only a concrete, material object. For this reason we prefer to use the term, a *composite commodity* (a bundle of commodities) instead of the simple word commodity. It is this composite commodity which has to be compared to another composite commodity to see if both give the same utility to a buyer. If they do, the two composite commodities constitute the same commodity.

There is another point that is relevant to this discussion of the meaning of the word commodity. If of two commodities understood in the above sense one gives double the utility of the other, they can still be made to constitute the same commodity by doubling the unit of the second commodity. Two units of the second commodity (now treated as one unit) would then give the same utility as the first commodity. If a thing gives 70 per cent of the utility of another thing it can, as said earlier, be treated as the same commodity to the extent of 70%: by increasing the size of the unit of this commodity by 3/7 we can make it yield the same utility as the other commodity.

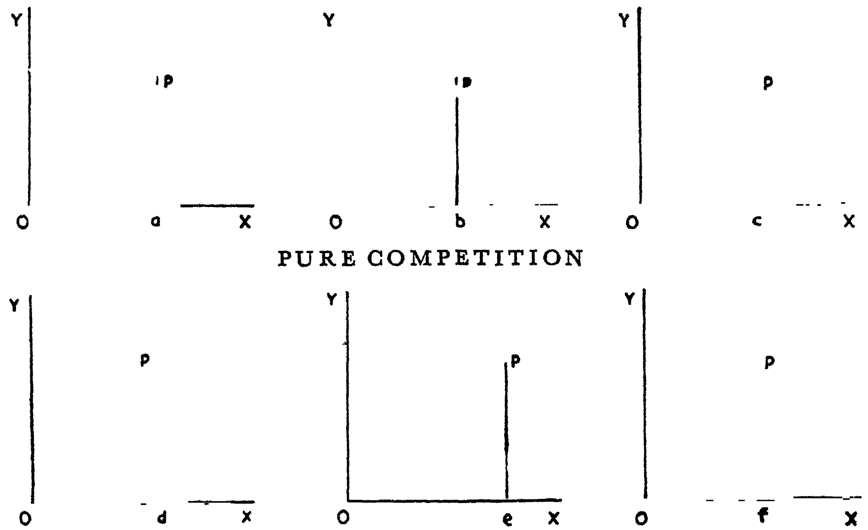
All this makes short work of the importance of the word commodity. For, things that are entirely different otherwise can be made to consti-

tute the same commodity as explained above. This is why, in defining competition, we made no specific mention of the word commodity.

PURE AND PERFECT COMPETITION

Unless otherwise stated, the word competition should imply pure competition, i.e., competition that is cent per cent competition. Milk should mean pure milk; buttter should mean pure butter; competition should mean pure competition. When, however, competition is mixed with (adulterated with) some alloy it ceases to be pure. The alloy can only be what is opposed to competition, i.e., monopoly. Hence, impure competition is a mixture of competition (pure) and monopoly (pure). Since competition is that market situation in which each seller has the ability to draw all the customers to himself and thus no seller is actually able to have any buyer (except, as explained above, when buyers are divisible), monopoly will imply that situation in which the seller is able to attract all the buyers to himself. Impure competition is, therefore, that situation in which, due to the admixture of a monopoly element, a seller is partly able and partly not able to capture buyers. When buyers are divisible, competitors will have equal numbers of buyers (see the argument given earlier) and impure competition will result in sellers having unequal numbers of buyers.

The difference between pure competition and impure competition can be illustrated with the help of diagrams (Diagram 15.1).



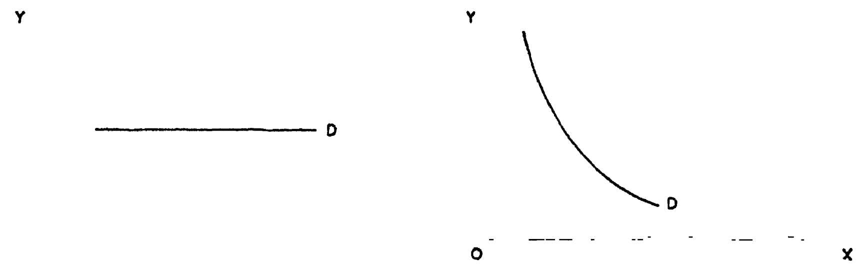
IMPURE COMPETITION

DIAGRAM 15.1

In all the graphs price is shown on the vertical axis and amount sold on the horizontal axis. The cases of three sellers are illustrated where there is pure competition and of three sellers where there is impure competition. In the case of pure competition the sales of all the sellers at the given price are equal ($Oa = Ob = Oc$) and the price $pa = pb = pc$. In the case of impure competition the sales are unequal at the price $pd = pe = pf$.

Certain conclusions can be drawn from the above facts. In the case of pure competition, a seller cannot sell more than his rival at the price ruling in the market. In the case of impure competition, a seller is able to sell more than his rival (the second and third sellers are able to sell more than the first seller) at the given price. That means, in other words, that a seller, if he chooses to sell the same amount as his rival, can charge a higher price. In the case of pure competition, therefore, a seller's demand curve is horizontal while in the case of impure competition it is slanting (negatively inclined). In the case of impure competition a seller can charge a price that is higher or lower than the price charged by others and yet his sales will neither shrink to zero nor increase to infinity.

Diagram 15.2 shows the demand curves for the product of a seller when there is pure competition and when there is impure competition.



Pure competition   Impure competition

DIAGRAM 15.2

The diagram of pure competition has a horizontal demand curve showing that, if the price is raised, there is no demand and, if it is lowered, howsoever slightly, there is infinitely great demand. Of course, the seller cannot meet such a huge demand. In the case of impure competition, the demand curve is sloping downward (it is negatively inclined), showing that if the price is raised there still is some demand and if the price is lowered the demand increases by a finite amount. It might be mentioned that, as impure competition is a mixture of (pure) competition and (pure) monopoly, the sloping demand curve itself should be composed of horizontal and vertical curves. We can imagine (and mathematically show) the sloping demand curve as made up of small horizontal and small vertical bits.

So far we have talked of pure and impure competition. We can now

make a distinction between impure and imperfect competition or between pure and perfect competition. This distinction owes much to Professor E. Chamberlin whose monumental book and articles on monopolistic competition have thrown a flood of light on the problems that we are discussing here.

According to Chamberlin, competition is pure when it is free from the monopoly element. We have used the term pure competition in this very sense above. The appropriate demand for the product of a single seller in the case of pure competition is, therefore, horizontal, as shown in the diagram above. With such a demand curve, the seller has no control over the price, in the sense that he cannot affect it by his individual decision to offer a smaller or larger amount for sale. But the situation may be such as to yield profit (excess of income over cost) which may or may not last long. It is also possible for a seller to incur a loss when his demand curve is horizontal. Whether he makes a profit or suffers a loss would depend on the position of his supply (cost) curve. According to us, as argued earlier, in the case of pure competition (that is logically pure and is also competition in the real sense) the sellers should have the same number of buyers. But let us leave that logically correct meaning of competition to the purest of pure economists.

In the case of perfect competition, according to Chamberlin, there would be no profit and no loss—price would just equal the average cost of production. Thus, the demand curve would not only be horizontal, it would also be tangential to the average cost curve. Chamberlin, however, leaves it to us to define perfect competition as we like. The definition would depend on our idea of perfection. However, we may note the fact that, if competition becomes pure in all industries, in all productive enterprises it would automatically become perfect too. The existence of profit (or loss) is ultimately due to the impurity of competition somewhere. Mobility of productive resources is one of the conditions for non-existence of profits and losses and that is also the condition for purity of competition. That is why it is possible to maintain that the condition for perfection of competition is the prevalence of pure competition everywhere in the economic system.

Competition and Number of Sellers

We have illustrated the case of competition (logically this word must connote pure competition) by taking two sellers, though we later passed on to the case where there are a larger number. In the literature on pure competition and impure competition we find mention of three broad categories: in one there is a single seller, in another there are a small number of sellers, and, in the third, there are a large number of sellers.

The first is referred to as the case of monopoly, the second as that of oligopoly (duopoly being a limiting case), and the third as pure competition of polypoly (usually called perfect competition).

To know how large the number of sellers is one must know what commodity is under consideration or which buyers are being catered to. For, a seller is a seller because he is selling a commodity or because he is catering to the needs of buyers. But a buyer is a buyer of a commodity and so, ultimately, one has to know the commodity that is traded in order to determine the number of competing sellers.

But, as we saw earlier, almost everything that is bought and sold constitutes the same commodity. For, in the final analysis, what a seller sells is utility and everything that is sold has utility. By varying the size of the unit of a thing that is sold we can make each unit of each object yield the same utility and thereby become the same commodity. That being done, the number of sellers becomes coextensive with the number of people that are sellers in any sense of the term. There is then only one answer to the question: How many sellers are there? The number of sellers is very very large. Cases of duopoly or oligopoly go by the board.

If we have to salvage the theory of oligopoly from this quandary we must take a more practical point of view and say, as Professor Joan Robinson says, that all those things that have more or less the same physical form and go perhaps by the same name must be regarded as constituting the same commodity. We would then have something to call by the name of oligopoly and be able to proceed to arrive at a solution of the price problem. But then we must be prepared to sacrifice some scientific precision: we have to pay the price for everything that we buy.

Oligopoly

When the sellers of a commodity are neither very many nor very few, they are said to constitute a case of oligopoly. Having taken a practical point of view (or a layman's point of view) in regard to the meaning of the word commodity, we have now to make a distinction between a small number and a large number. Oligopoly implies a small number of sellers and though it should not, as we said above, include the case of a very small number, it has been allowed in economic literature to cover the case of even two sellers, again, to overcome practical difficulties.

The first and the important point is to draw a line between a small number and a large number of sellers. The distinction between a large number and a small number of sellers hinges on the effect of a change in output or price made by one seller on the decisions of other sellers in regard to their production or sale. If, for instance, a fall in the price charged by a seller induces other sellers to follow suit the number of

competing sellers is said to be small. When the other sellers are not *induced* to change their decision, the number is said to be large.

When there are only two sellers, it is quite obvious, a change in the behaviour of one induces a change in the behaviour of the other, either with the object of protecting his interest or with the object of harming the interest of the rival. In other words, a change in the behaviour of one seller induces a defensive or offensive action on the part of the other. The case of a duopoly falls in the category of what we have called (in a later chapter on Closed and Open Systems) an open system. And as we have shown there, no equilibrium can be reached so long as the system continues to be open.

So let us take up the case of duopoly which is unmistakably a species of oligopoly. We shall examine the solutions of A. Cournot and F. Y. Edgeworth which have become classic in economic literature. Suppose there are two sellers of a commodity, and their object is to maximise their net money income. This is the traditional assumption of maximising behaviour. We shall proceed by taking cost and demand functions, cost function being a quadratic equation and demand function being a linear equation. This procedure has now become a commonplace in economic literature. For the sake of simplicity, we shall assume that both the sellers have identical cost functions and that they cater to the needs of a common group of buyers. These assumptions very well suit our conception of competition. Anyway, let us proceed with the mathematical solution of the problem. Let the total cost functions of the duopolists be $T.C. = Ax^2 + Bx + C.$ and let the demand curve which they face be $p = b - ax$. By symmetry, we can at once jump to the conclusion that the equilibrium solution will yield equal outputs for the duopolists. Let us, however, take their outputs to be x_1 and x_2. The total output would then be $x_1 + x_2$.

The net income of the first seller would then be given by

$$\begin{aligned} \text{N.I.} &= px_1 - Ax_1^2 - Bx_1 - c \\ &= x_1\{b - a(x_1 + x_2)\} - Ax_1^2 - Bx_1 - C \end{aligned}$$

To maximise this, its first differential with respect to x_1 has to be zero.

$$\frac{\partial(\text{N.I.})}{\partial x_1} = 0 = b - 2ax_1 - ax_2 - ax_1\frac{dx_2}{dx_1} - 2Ax_1 - B \qquad \ldots(1)$$

Similarly for the second seller we would get the condition

$$\frac{\partial(\text{N.I.})}{\partial x_2} = 0 = b - 2ax_2 - ax_1 - ax_2\frac{dx_1}{dx_2} - 2Ax_2 - B \qquad \ldots(2)$$

These equations give the conditions for the maximum net income for each seller. If we solve these equations simultaneously we can get the value of x_1 in terms of x_2 or *vice versa*. Solving simultaneously implies that the two sellers are at the same time in equilibrium, with their net in-

comes raised as high as possible. The values of outputs that we would get are those that would be arrived at if and when the two sellers are simultaneously in equilibrium. It is not necessary that such a position should be reached in actual practice at any finite interval of time.

As a matter of fact, we start with different values of the outputs of the duopolists. The two sellers go on reacting to the moves of each other—the system is open and remains open and, as said above, equilibrium cannot be attained till, by accident or otherwise, the sellers stop reacting to each other's decisions in regard to the amount to be sold.

In solving the equations (1) and (2) simultaneously we have to assign certain finite values to dx_1/dx_2 and dx_2/dx_1. The former of these stands for the change in the output of the first seller that the second seller thinks would be *induced* by a small change in the output of the second seller. Likewise, the second expression stands for the change that the first seller thinks would be induced in the output of the second seller by a small change in his output.

Cournot, in order to be able to solve the equations simultaneously put $dx_1/dx_2 = dx_2/dx_1 = 0$. This means that it is assumed by each seller that the rival will not retaliate by changing his output when he changes his own output. This is hardly a realistic assumption. As a matter of fact, the definition of a *small number* of sellers is that the change in the output or the price of one affects the output or the price of the other. But if we have to have a mathematical solution we must sacrifice the essence of duopoly. If we do not, we cannot get a solution and this is what we have emphasised above and in the chapter on Closed and Open Systems. There can be no equilibrium so long as the system continues to be open.

So, putting the differentials of one output in terms of the other equal to zero, we can proceed to solve the equations for the relative values of x_1 and x_2. Here we can take advantage of symmetry and conclude that in equilibrium the outputs of the two sellers would be equal. We can, therefore, put x_2 equal to x_1 and thus eliminate x_2. The solution would then yield the value of x_1 in terms b, B, a, and A. And that would, therefore, be the value also of x_2. This value comes out to be

$$x_1 = x_2 = \frac{b - B}{3a + 2A}$$

The maximising equations can also be solved simultaneously by assigning some finite numerical values to dx_1/dx_2 and dx_2/dx_1. If we do this we can meet the above objection, namely, that duopoly becomes a closed system, thus destroying the very essence of a *small number* of sellers. But can we really assign any finite numerical values to these differentials? For, as competition proceeds, the values will go on changing. A.C. Pigou rightly said that duopoly has no solution without assumptions and that no realistic assumptions can be made.

Let us examine Edgeworth's solution of the duopoly problem. Suppose, as before, that the two sellers have identical supply curves and that they have to face a common group of buyers. Let the supply and demand function be the same as before

$$\text{N.I.} = px_1 - Ax_1^2 - Bx_1 - C \text{ for seller No. 1}$$

$$\frac{\partial(\text{N.I.})}{\partial x_1} = p + x_1 \frac{dp}{dx_1} - 2Ax_1 - B = 0 \text{ (maximising condition)}$$

Similarly,

$$\frac{\partial(\text{N.I.})}{\partial x_2} = p + x_2 \frac{dp}{dx_2} - 2Ax_2 - B = 0 \text{ for seller No. 2}$$

We can solve simultaneously the above equations that give us the maximising conditions. Simultaneous solution will yield values of outputs which, if and when attained, would put an end to further reactions to each other's changes in output decisions. But here also, for a mathematical solution of the equations, we have to assign some finite numerical values to dp/dx_1 and dp/dx_2. Edgeworth assigned the value zero to them and thus simplified the solution. In the language of economics, this means that a duopolist thinks that when he fixes his output or changes it in order to maximise his net income, the rival will let his price remain unchanged. Anyway, the equations are solved simultaneously and, taking advantage of symmetry x_1 is put as equal to x_2 and an easy solution is obtained. The outputs turn out to be

$$x_1 = x_2 = \frac{b - B}{2a + 2A}$$

It may be noted that this is the equilibrium solution, meaning thereby that, if by chance the two duopolists happen to offer this amount, then, under the assumption made, they would make no further change in their outputs and, therefore, in their prices. Otherwise, as Edgeworth maintained, while competition goes on, the price would continually vary, oscillating over a range in between the monopoly price and the competitive price. Again, to say that a stable equilibrium price would be reached in the case of duopoly would be wrong. For, duopoly is an open system and no equilibrium is possible. The oscillating price suggested by Edgeworth is a more realistic and scientifically correct solution. But the assumption that each duopolist thinks that the price of the other will not change is unrealistic.

The problem of duopoly or oligopoly has been solved in other ways also. We can make different assumptions about the behaviour of sellers and get different solutions. We may assume, for example, that one seller assumes the role of a price leader, or is believed by others to assume such a role, and that others follow the lead given by him. We can also assume that the competing sellers, sooner or later, come to some open or

tacit agreement in regard to price or agree to divide the market among themselves. We might also make assumptions in regard to the reactions of a seller to the moves of other sellers, reactions that are variable and not constant. One such assumption that has some fascination for students of economics is that a seller's demand curve has a kink. This means that there is a more or less sudden change in the slope of the demand curve.

Diagram 15.3 shows such a kinked demand curve. One gets such a curve when it is assumed that one's rival will lower his price when one lowers one's own price but that such a reaction would not be there when one raises one's price. This may sometimes happen but it is not necessary that it should happen always. At times it becomes more profitable to raise the price a little when the rival charges a higher price. The wisdom of doing that would depend on the elasticity of demand.

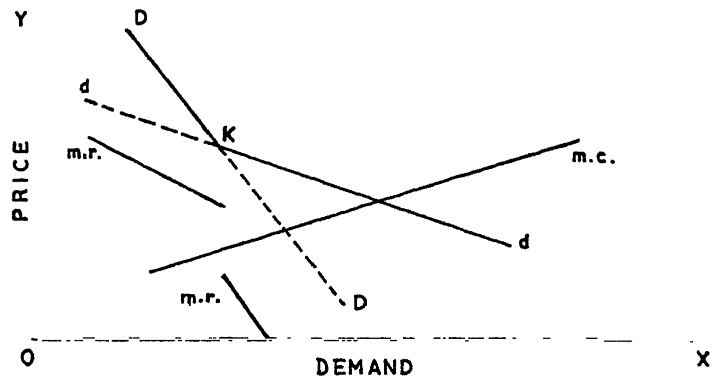


DIAGRAM 15.3

Cournot assumed that each duopolist thinks that the rival will keep his output unchanged while Edgeworth assumed that each duopolist thinks that his rival will keep his price unchanged. Thus, according to Cournot, the demand curve would have a steeper slope and, according to Edgeworth, a gentler slope. In Diagram 15.3 according to Cournot's assumption, the demand curve would be somewhat like *DD* while, according to Edgeworth's assumption, it would be like *dd*. Now, we can combine the two to get the kinked demand curve. Let us suppose that a fall in price by one is followed by a fall in the price of the rival while a rise of price is not followed by a rise. The resulting demand curve would then be composed of parts of the two curves as shown by the discontinuous curve having a kink at the point *K*.

A demand curve is an average revenue curve and it has its marginal revenue counterpart. The marginal revenue curve corresponding to the broken or kinked curve has two components — the marginal revenue

curve falls abruptly when it reaches a point just below the kink. The marginal revenue curve is said to be discontinuous. If the marginal cost curve be as shown in the diagram it would not cut the marginal revenue curve. If net income is maximum when the two curves intersect, the diagram yields no finite point of maximum net income. The price, we are told, is indeterminate. This statement should be made with caution. For, there will still be some price which will yield the best result. This can be found by comparing total revenue with total cost.

Place of Selling Costs

It is said that in the case of perfect (pure) competition selling costs are not necessary. The reason given is that an individual seller can sell as much as he likes by merely reducing his price slightly below that charged by others. So that this may be possible the individual seller's supply should constitute a very small part of the total supply. Hence, the conclusion further drawn is that, when there are a large number of sellers, selling costs are unnecessary.

According to our definition, there can be pure competition even when there are only two sellers, provided each is able to draw all the customers to himself. A customer is attracted or repelled by two factors: one is the commodity offered to him and the other is the price demanded for it. Assuming the commodity is the same in the sense in which we have defined it earlier, price is the only thing that attracts or repels a customer. If a seller has to be able to attract customers to himself he should be able also to vary the price to suit the circumstances. And if by varying the price (say, by reducing it) customers are attracted, it means that they are attached only to the price and not to the seller for any reason whatsoever. But a customer, as a customer, is never attached to the seller (and is only attached to the price) when the commodity sold by all the sellers is the same. And since the commodity can be and, therefore, is the same, the only condition for perfect (pure) competition is that buyers should not be attached to sellers and, as said above, they never can be attached to them. If this argument is accepted, competition must always be perfect and cannot be otherwise. If, however, the possibility of impure competition has to be admitted, the word commodity must be understood in its ordinary sense and things that look alike must be allowed to constitute the same commodity and those that do not should be regarded as different commodities.

When the word commodity is thus defined there would be some attachment to particular sellers in spite of their all offering the same commodity to buyers. To make competition pure (or perfect as it is often called) the other subjective elements that go together with a commodity

and are thus offered to buyers should in their totality make the same appeal to buyers.

It is here that selling costs make their appearance. Things that otherwise look alike can be made to make a greater appeal to a buyer if they are effectively advertised. And, likewise, things that appear different can be made to be equally attractive to buyers when suitably advertised. And the selling cost thus incurred would form a part of the cost of production of the commodity.

Monopoly

Etymologically, the word monopoly means one seller. But when one says that there is only one seller in the case of monopoly one has to specify the commodity and, perhaps, also the area over which he operates. These difficulties can be circumvented by defining monopoly in a different way. Let us say that a monopolist is one who can charge any price he likes for the commodity he sells. If the word commodity is defined in its widest sense, all the things that are sold anywhere would constitute one and the same commodity. Theoretically, then, a monoplist would be the only seller if he has to be able to charge any price he chooses to. All buyers of all the objects (for they now constitute the same commodity that the monopolist sells) would become the customers of the monopolist. For, he can satisfy their wants, his commodity being theoretically the same as any other. If such a monopolist can be imagined to exist, he would oust all other sellers and remain the only one for all buyers to go to.

The limited resources of buyers set, however, certain limits to the power of a monopolist to raise the price of his commodity. He would, therefore, be a monopolist over a limited range of prices. The demand curve for his product would be vertical but would extend to a height that has a maximum limit. Just as it is hard to find a case of absolutely pure competition (where the demand curve is horizontal *ad infinitum*) it is hard also to find a case of pure monopoly. What exists in this material, phenomenal world is a mixture of impure and imperfect things. The moment a thing becomes perfect it ceases to exist as such. All things exist in their imperfection. If we picked up a seller at random, we would find the demand curve for his product to be neither indefinitely horizontal nor indefinitely vertical. So, coming down from theory to practice, we find the demand curve to be a sloping one, negatively inclined. And, as we have shown in another chapter, such a curve is formed of horizontal and vertical bits — it is a mixture of pure competition and pure monopoly that we find in practice. Now, if that is the case, every seller would appear, to some extent, to be a monopolist and, to some

extent, a member of a competing group. Professor Joan Robinson, for this reason, maintained that every seller was a monopolist. According to Professor E. Chamberlin (shall we say for the same reason!) every seller belongs to a group of monopolistic competitors.

One can find a derivative to measure the monopoly element in a situation, or the monopolistic power of a seller. Since the verticality of a demand curve is due to the monopolistic power of a seller, we can measure the slope of a curve to determine the degree of monopoly in a given situation. If, for example, the demand curve is represented by the equation $p = f(x)$, where p is the price and x the demand, the expression dp/dx would show, by its mathematical value, the verticality of the demand curve. It could, therefore, be said that the greater the arithmatical value of dp/dx, the greater is the degree of monopoly power that the seller commands. In the case of theoretically absolute monopoly, dp/dx is infinity and, in the case of pure competition, it is zero.

The traditional way of determining the price that a monopolist would charge is via the equation of marginal cost and marginal revenue. But the demand curve in such a case is negatively inclined because the monopolist does not have a full and thorough grasp on the entire body of buyers. This can be explained as being due to the fact that, either the buyers who have the monopolist as the only source from which to buy have other commodities that compete with the one in question for a claim on the purchasing power of the buyers, or there are other sellers of the same commodity to whom the buyers have access. If the negative slope of the demand curve is due to the latter reason, it is most unlikely that the monopolist will have a curve that remains unchanged. For, in such a case, the system is an open one and the action of one seller induces reactions from other sellers. The determination of price would then not be the simple affair that it is taken to be. The monopolist does not know how his rivals will react to his attempts to maximise his net income nor do the rivals know how the monopolist will react to changes in their own price and output decisions. The solution of the price-problem, then, must depend on the analysis adopted by Professor Chamberlin and Professor Triffin.

If, however, the slope of the demand curve is simply due to the fact that there are other commodities that make a claim on the income of buyers and thus compete with what the monopolist sells, the system becomes a closed one. Maximisation of net income is then effected by the monopolist in the usual way, by equating marginal cost with marginal revenue. This, according to Professor Joan Robinson, becomes the most typical and the most general case of buying and selling—monopoly (thus understood) swallows up all cases of the buyers-sellers relationship.

CHAPTER 16

Functional Finance

Non-Functional Finance

This is rather an odd phrase but it is used with the object of distinguishing the old views on public finance from the new ones. In the past, public finance was concerned with the principles that governed the raising and spending of money that was needed for the performance of public activities which consisted mainly in preserving order within the country and defending it against foreign aggression. The other duty of the state was to look after public health. For the performance of these duties, money was needed and it was raised from the people in the best possible way and then spent in the most efficient manner. Public revenue, public expenditure, public debt and financial organisation were the traditional heads under which all the principles of public finance were grouped.

The duties of the State increased in number as it was realised that, besides the above, there were many other things that it could do, i.e., when it came to be realised that a number of other wants of the people could be more effectively and efficiently satisfied if they could pool their resources for the purpose and adopted co-operation rather than competition as a means of satisfying their wants. However, even then public finance confined its scope to the study merely of the principles that governed the getting and spending of money that was needed for the discharge of public duties. Public finance, thus limited in scope, appears to us today as non-functional in its nature. But the line that we can draw between functional and non-functional finance is rather a broad one.

Functional and Non-Functional Finance

Today it is realised that the objectives of public finance can very legitimately be widened. The State can do a number of other things besides merely protecting the people against internal and external violence and keeping them physically and mentally fit and healthy. Take for example the objective of fuller employment. If there is a good deal of unemployment the government can do something to reduce it. If the unemployment is found to be due to absence of efficient means of communication, the appropriate remedy would be to build good roads. If unemployment can be reduced by helping private enterprises to raise their productive efficiency, that can also be done by supplying them the means they need. Such measures would legitimately fall within the scope of traditional, non-functional finance. In the same way, if we want to reduce inequality of wealth or income it can be done by adopting means that would have a place in the old system of public finance. For instance, by legislation, the wages of labourers can be raised and thus, to some extent, greater equality of earnings secured.

Functional finance has the same objectives, at least some of the objectives are common. But the way of attaining them is different. For example, to increase employment the government pumps extra purchasing power into the economy. If this purchasing power goes into the pockets of those who would spend it rather than hoard it, production and employment would increase through the multiplier effect. The difference between functional and non-functional, traditional finance is not so much one of objectives aimed at as of the means adopted for that purpose. In what follows these points will be fully explained.

Objectives of Modern Public Finance

The most fundamental objective into which all other objectives can be shown to resolve themselves is maximum social welfare. No objection can be taken to this being accepted as the ultimate end of all activities of the government that fall within the scope of public finance. But the administrator does not know in what social welfare—the welfare of the people—consists and how it is to be estimated with a view to the determination of its maximum level. An individual knows or feels his welfare and perhaps can say whether it is maximum in certain given conditions. But the administrator does not know what is the welfare of the collective body of people and whether it is maximum or not. As a matter of fact, social welfare has been the bugbear of welfare economists. We have

explained the difficulty involved in attempting to apply the criteria of optimality. We do not need to say anything more about it here.

Coming from the ultimate to the mediate (or immediate) ends of public finance, we find a large number that demand our attention. We can and need to mention only some of them here. Starting with an objective, measurable entity, we can mention income as one of the objectives. Naturally, we would aim at maximum possible income, maximum possible for the average member of the state. In familiar language, our objective may be maximum per capita income. If all that should be included in income is included, no serious objection can be taken to this objective. Other things remaining the same, it appears to be most desirable to aim at as high a per capita income as possible.

But other things may not and do not remain the same and so we have to consider other objectives also. Income should fluctuate as little as possible from time to time. If it continuously or continually increases so much the better. But it is better to have a stable income than an unstable one. With instability, there is uncertainty and uncertainty is never preferred to certainty. Even when there is certainty about unstable income, it is not to be preferred to stable income. This follows from the fact that equality of income, even as between one period of time and another, is better than inequality. Just as we feel (all may not feel that way, however) that there should be equality of income as between two individuals, so also we feel that there should be equality of income as between one period of time and another.

This leads us directly to the associated objective of equal or equitable distribution of income. It may be more desirable to have a smaller income equally distributed than a larger income unequally distributed. But perhaps there is no proof of this proposition as distribution involves a value judgment. However, most of us would recognise reduction of inequality of income as a legitimate objective.

After income, we may take up employment. Though income is a more fundamental objective than employment, for the reason that employment is a means to income, yet employment is also a necessary thing to aim at. Unemployment, besides depriving one of monetary income, exercises a demoralising influence on one. If national income can be maximised with some people unemployed, it is yet desirable to adopt measures to bring about fuller employment so that the people may enjoy life more fully, with the consciousness that they are not living on public charity. It is also desirable to have non-human factors of production fully employed, but that would be only with the object of maximising income. For, material things do not have a mind or a heart and would not, therefore, suffer a loss of psychic income in the case of their being unemployed. It is most important, thus, to use the tool of public

finance to bring about such changes in the economy of a country as would lead to a reduction of unemployment of human factors of production.

Why is There Underemployment?

If left to themselves, economic forces, it was believed, would so adjust themselves as to bring about a state of full employment of all resources. This, however, does not happen for a variety of reasons. To start with, let us note that even when one factor is fully employed the other factors may remain partially unemployed for the reason that the coefficients of production are not sufficiently elastic. There are technical limits to changes in the proportion in which factors of production can be combined. Hence, even when capital is fully employed labour may remain underemployed.

There are other reasons also for underemployment, but we shall concentrate our attention on only one which is amenable to treatment with the tools of public finance. Classical economists, by and large, believed in what is known as Say's Law of the Market according to which there can be no general unemployment as supply always creates its own demand. Other things being the same, supply would certainly create its own demand. Take, for example, the simple case of a farmer employing a labourer. When corn is grown, wages are earned by the labourer and so he gets the purchasing power needed to buy the corn produced. The supply of corn is instrumental in creating purchasing power and, therefore, also the demand for corn. General unemployment, J. B. Say said, is for this reason not to be feared. This reasoning assumes that there will be neither time lags nor leakages, that income earned would be spent without delay in buying the goods that are instrumental in producing it and that no part of income would be frozen, i.e., there would be no hoarding of money.

Keynes argued to prove that it is possible for an economy to reach a position of equilibrium even when some factors of production are not fully employed. If this is so, it must be because there is some obstacle to the operation of Say's Law of the Market. This obstacle is the hoarding of money; at any rate this is the most important factor that brings about the state of underemployment equilibrium. The income generated by the productive machinery is not fully spent in buying what is produced or in replacing worn-out equipment. What is needed is that the flow of money from the producers to the consumers and savers and back to the producers must be kept up. If that does not happen, production will have to be cut down, causing unemployment of some units of labour and other factors. And that does not happen when there is hoarding of money, hoarding of purchasing power. Those who earn income should

spend the whole of it. We are using the word *spend* in its economic sense. Money should be used to buy consumption goods or producers' goods. The entire income should be spent (used in buying consumption goods) or saved and, ultimately, invested. If the saved money is not invested, it remains hoarded. It may be asked why savings are not fully invested, or in other words, why a part of money remains hoarded.

Investment is needed to produce goods and if they cannot be sold investment is wasted. Hence, all the money that is saved will be invested only if the entire output can be sold at a remunerative price. If the (not saved) money spent on consumption goods is insufficient to buy them at remunerative prices, some stock of goods would remain unsold and production would be cut down and unemployment would follow.

What is required, therefore, is that only a certain proportion of money earned should be saved. As will be seen later on, the ratio of expenditure to saving should correspond to that of output of goods to investment. To sum up, we can, following Keynes, say that unemployment or underemployment is due to too much of saving and too little of spending resulting in deficiency of effective demand for goods. It may be mentioned in passing that there can be other causes of unemployment also but we are here concerned with this particular cause which is important in economies that are fairly developed. We shall explain below how public finance activities or financial measures can tackle this problem of deficiency of effective demand.

CAUSE OF DEFICIENT DEMAND

The demand for goods becomes deficient when more than a certain amount is saved or less than a certain amount is spent. What proportion of income spent ensures sufficiency of demand will be explained later on. Here let us explain when or why less than a sufficient amount of income comes to be spent. Let us argue in terms of an individual income-earner. An individual has the option of spending the whole of his current income or a part of it. He has also the additional option of over-spending (borrowing being left out of account) by drawing upon his past savings, if any. In normal times, an individual always saves a part of his current income. What part he saves depends on a variety of factors. But we know that the richer the man, other things being the same, the greater is the proportion of income that he will save. A poor man saves a smaller part of his income than a rich man. This is because, for a poor man, the discounted marginal utility of saving is less than the marginal utility of spending as compared to a rich man, for whom it is relatively more.

If, therefore, a great proportion of total national income accrues to

comparatively rich people, a large part of it would be saved. The full-employment ratio of saving to spending is, therefore, exceeded when the bulk of national income is earned by rich people. Likewise, the full-employment ratio of saving to spending is not reached, i.e., less than the required amount is saved, when the bulk of the income goes into the hands of poor people. To rectify the state of an economy in which there is deficient demand we have, therefore, to devise means to put a larger proportion of national income into the hands of poor people. If, however, more than that amount goes to the poor, the economy witnesses boom conditions and, after full employment level is reached, prices begin to rise without a rise in real income. The cause of deficient demand is, therefore, the concentration of a large part of national income in the hands of those whose propensity to consume is low.

When Does Distribution of Income Get Tilted in Favour of the Rich?

In the study of this problem, agents of production are broadly (and unscientifically) divided into two classes, viz., wage-earners and rentiers. The former includes labourers, organisers and capitalists (lenders of money), while the latter includes only the producers. This, we have said, is an unscientific division in as much as the producers are not really rentiers and there is no separate class of people that can be called producers. Labourers, for instance, are also producers. But there is some justification in the use of the word rentiers because, when prices rise, the producers, that is, those who employ hired agents of production, get a surplus and, when prices fall, they suffer a loss. The gain or the loss (due to which the producers are called rentiers) is often temporary. For, wages and the earnings of other hired factors sooner or later rise as a consequence of rise of prices. Yet, for a short period, producers do enjoy a surplus income when prices rise and, if they keep on rising from time to time, they keep on enjoying surplus income.

If an economic system is prospering and national income is rising, the first sign of this that is generally noticed in the upward trend of prices. And when prices rise, the producers' income rises faster (more) than the incomes of other agents, the bulk of whom consists of wage-earners. Hence, during a period of prosperity and a rise of national income, purchasing power tends to shift from the relatively poor to the relatively rich.

It is, therefore, natural for the effective demand to become deficient sooner or later when national income is rising. But the rise of income may be a phase of the trade cycle in which it would be followed by a fall of income, or it may be a manifestation of the rising trend of economic

activities. In other words, there are two types of rise of income, one that is due to causes that produce trade cycles, the other that is due to the growth trend in the economic development of an economy.

EQUILIBRIUM WITH UNDEREMPLOYMENT

We have seen how underemployment of resources (and we are most concerned with the underemployment of labour) results from insufficient spending. If more than a certain amount is saved, all that has been produced for sale cannot be sold. Production then becomes unprofitable and has to be cut down, causing some unemployment. We have also seen that such a situation arises when purchasing power gets concentrated in the hands of the richer classes whose propensity to consume is relatively low. And such a concentration of income, if it is not there for any other reason, is bound to follow a rising trend of income. If the income is increasing because the economy is passing through the upswing phase of the trade cycle, the adverse effects of insufficient spending, otherwise caused, are neutralised by the rising tempo of economic activities that keeps on pumping in purchasing power into the hands of the people, the poor included. If, on the other hand, the concentration of income in the hands of the rich is caused by general growth of the economy as a result of new techniques of production, the adverse effects of insufficient spending are again counterbalanced by the continual injection of fresh purchasing power into the economic system.

Underemployment is, therefore, held in check in those circumstances in which the very forces that tend to depress the expenditure of money also lead to its increase. Hence, if there has to be underemployment, and at the same time equilibrium in an economy, depressed spending should not be the result of rising national income. This takes us to the question proper of underemployment equilibirum. If an economy settles down at a low level of economy, conditions become disquieting indeed. For, we can then describe the state of the economy as one in which it has meekly submitted itself to the phenomenon of underemployment. It becomes, in such a case, necessary to pull the economy out of the pit, to do something to make it realise that all is not lost and that better times can be ushered in.

It is here that public finance can come to the rescue of an economy. Let us mention first that such a situation of underemployment equilibrium arises when insufficient spending, leading to increased saving, is ultimately balanced by investment. Equilibrium requires the balance of saving and investment. Due to deficient effective demand, investment decreases; due to decrease of investment, income decreases; and, due to decrease of income saving again decreases. Ultimately, saving and

investment become equal, at a low level of course. The economy then stagnates, people are unemployed though these are willing to work, but it does not pay the employers to employ them.

The remedies that come to our mind to stimulate the system to enable it to employ more resources are, first, reduction of wages, and, second, lowering of the rate of interest. It is true that, if wages could be reduced without inducing other neutralising changes in the economy, it would become possible to induce employers to increase production and thereby increase employment. In the same way, if the rate of interest could be lowered and made attractive to producers, investment would increase and, along with that, employment. But for certain reasons this is not possible. The good effects of such changes are neutralised by other changes that take place in the economy. Let us explain these points in the section that follows.

Lower Cost of Production Fails to Increase Employment

Since it is the producers who employ workers, it should be made attractive to them to increase the employment of labourers. This can be done in the most natural way by lowering the cost of production. This cost can be lowered by reducing wages or by lowering the rate of interest. Both these measures fail, however, to induce producers to employ more men.

Lower interest fails to induce producers to increase production because of what is called the *liquidity trap*. The cost of production no doubt falls, but it does not fall sufficiently to make additional investment attractive. Moreover, people find it more profitable to hold money in cash than to invest it at a low rate of interest. When the preference for liquidity is great (as it generally is when times are not prosperous), people would rather go in for liquid assets than for securities. For this reason, money is trapped in the liquidity trap and is not able to enter the investment market. Much cannot, therefore, be expected from this measure until the psychology of producers changes, making them optimistic about the future.

The rate of interest can be lowered by the monetary authorities by increasing the supply of money. They have to depend on the laws of economics to effect a fall in the rate of interest. Interest, like any other price, can fall only when the supply increases, demand remaining the same, or when there is a relative increase of the supply of money. Unfortunately, what happens is that, when the supply of money increases, the demand for it also increases, not because the supply has increased but because of the operation of those very forces which make it necessary to

increase the supply of money. We are referring to a depression; for then economic activity is at a low ebb and we want to give it a fillip by bringing down the rate of interest. But in these very circumstances, people are unwilling to increase their investment and prefer instead to increase their cash balances. There is preference for liquidity over that for investment. Due to this preference, the demand for money rises—people want liquid money and that is the meaning of increase of the demand for money—and the rate of interest is prevented from falling or falling adequately. The liquidity trap operates.

Reduction of wages also fails because, although the cost of production falls, other things, that could by their change neutralise the good effect of lower wages, do not remain the same. While low wages shift the cost curve to a lower level, they shift the demand curve also to a lower level. It is assumed by those who advocate reduction of wages that other things would remain unaffected. But, as Keynes argued, low wages might (if the elasticity of demand for labour is not high enough) reduce the demand for goods by decreasing the total earnings of wage-earners. The benefit of low cost of production is shared by all those who buy the goods produced while the loss caused by it (i.e., by low wages) is suffered only by wage-earners. For this reason, non-wage-earners, by their increased demand for goods, might be able to offset to some extent the adverse effect on production and employment of the lower earnings of labourers. But the over-all effect of wage reduction would be to reduce employment rather than increase it. As we have said above, it is the wage-elasticity of demand for labour that would determine the final effect. It might be noted however that lower wage is not a dependable remedy for underemployment. Since these measures cannot be relied upon, we have to take recourse to fiscal measures.

The object of fiscal measures is directly to provide employment for labour. Of course, we know that labour alone can never produce anything and so, with the employment of labour, there goes necessarily the employment of other factors. Anyway, to employ labourers the government needs money. This finance has to come from taxes, loans or created money. Because of this, the direct employment of labourers by government is called a fiscal measure for reducing unemployment. This type of measure can be adopted by private individuals also, but they would not have the courage to do so because of the depressed condition of the market and the consequent pessimism that characterises the mental attitude of producers in such a situation. Further, we may note that a private individual cannot tax the citizens of a country, nor can he create money. For all these reasons, the adoption of fiscal measures by individuals is never considered.

We shall not take up the question of the relative importance or efficacy of the different ways of financing the projects through which a govern-

ment attempts to tackle the problem of underemployment equilibrium. Let us instead see, first, how fiscal measures lead to employment of labour. Suppose the government spends (which means invests) a crore of rupees in building a bridge across a river. If this money is spent within a year it will go into the pockets of the factors of production engaged directly or indirectly in the construction of the bridge. This money would then be spent on goods (we always include services in goods) that the economy is currently producing. The first result of this would be to raise the price level. To what extent it will rise will depend on a variety of factors on the demand and supply sides. If a large part of the money goes into the pockets of poor people, or at any rate those whose propensity to consume is high, the price level would rise considerably. And then the prices of wage-goods or those goods that are consumed by those people would rise more than the prices of other goods. The extent of this rise would also depend on how the producers and sellers respond to increased demand. If they allow their inventory to get deflated prices would rise less. There are other factors also on which the extent of price rise would depend.

However, with the injection of a crore of rupees into the system prices of some goods are bound to rise more or less immediately and those of others sympathetically, after a time lag. This is the primary effect of government or public investment. The investment not only provides employment to labourers in the first round, but also, by increasing the demand for currently produced goods, leads to a further employment of labourers in the industries producing those goods in the second round. This is the secondary effect of public investment. In this way the employment of labour and national income go on increasing. After the secondary effect, there is a tertiary effect when the people engaged in the production of goods to meet the increased demand from labourers originally employed by the government begin to spend *their* income. Income is first earned and then spent on goods already available, thereby raising the profit margin of producers. This leads to further production through further employment, raising the national income thereby. The time lag in the process explained above leads to increase of income and employment by stages. The income and employment keep on increasing till, ultimately, factors of production become scarce. There cannot be full employment of all the factors at the same time nor even at any time due to the inelasticity of coefficients of production after a certain limit. However, our purpose is served if unemployment is reduced to the minimum. This process of initial dose of investment and employment giving rise, by stages, to greater and greater investment and employment is called the multiplier process.

BALANCED AND UNBALANCED BUDGET MULTIPLIER

We have seen how employment can be increased by public investment through the multiplier effect. Since underemployment is due to insufficient spending (when it is due to that), what is required is to increase spending. And that is done by injecting extra purchasing power into the economic system. That has the effect of raising prices and increasing the profits of producers. Prices rise because there is more money and an unaltered supply of goods. Production, therefore, increases and, along with that, employment. The primary effect is followed by a secondary effect and then the tertiary, and so on, till full employment of labour or other factors is reached.

Now, if this has to happen, an extra amount of purchasing power must be pumped into the economic system. The best way to do this is to print notes and use them to pay the factors of production employed in new projects. This is called deficit financing in public finance. In the case of deficit financing, the money that is put into circulation does not come out of the pockets of the people. The other way to finance government projects is to get the money through taxation. But in that case the purchasing power does not increase, as what is given to some people is taken out of the pockets of other people. It would appear, therefore, that employment cannot be increased by public investment if it is financed by taxation, i.e., if there is a balanced budget. We shall see in what follows that even when money is taken out of the pockets of some people and put into those of others (i.e., even in the case of a balanced budget), the total purchasing power increases to some extent.

However, even when public undertakings are financed by created money, i.e., by printing notes, employment will not increase satisfactorily unless certain conditions are fulfilled. For instance, the multiplier effect would not be realised fully if the supply of money does not easily increase. In the first instance, the government prints notes and finances its projects. For the primary increase of employment that is enough. But if there are to be subsequent effects, that is, secondary effects, tertiary effects and so on, the monetary or banking authorities must allow the supply of money to increase in response to the demand for it. If the banking authorities do not grant credit to the extent to which it is needed (when there are secondary and subsequent increases of demand for goods) neither production nor employment would increase according to the operation of the multiplier.

Further, the propensity to consume should not decrease. If the people do not spend the money they earn or not as much as they were spending before, the demand for goods will not increase as expected and, con-

sequently, the rate of increase of production and employment will be retarded. There are other conditions also that should be fulfilled, otherwise the multiplier will not operate freely. But we are concerned here with only these two impediments to rapid and satisfactory increase of employment.

The Multiplier—Its Evaluation

Let us now explain how the value of the multiplier is calculated. Let it be noted that an initial increase of income effected by the printing of notes to finance government undertakings produces, ultimately, a multiplied income. If the income finally increases to twice as much as in the beginning we say that the value of the multiplier is 2. If it increases to four times the multiplier is said to be 4. If income increases four times, that is, if the initial government investment of one crore of rupees increases income to the extent of four crores of rupees, employment would also increase four times provided the production function is linear and homogeneous. In simple words, if constant returns to scale prevail employment and income would increase to the same extent. The income multiplier and employer multiplier would be equal.

Let us see how the income multiplier is measured. We represent income by Y, increment of income by ΔY, investment by I, increment of investment by ΔI, saving by S and increment of saving by ΔS. If due to the investment of ΔI income increases by ΔY, the multiplier is equal to $\Delta Y/\Delta I$. Let us call the multiplier K.

In equilibrium—in dynamic equilibrium—savings are equal to investment and, therefore, we can substitute ΔS for ΔI and get $K = \Delta Y/\Delta S$ or $K = \Delta Y/(\Delta Y - \Delta C)$ where ΔC stands for increment of consumption. With this substitution we get

$$K = \frac{1}{1 - \Delta C/\Delta Y} = \frac{1}{\text{1-marginal propensity to consume}}$$

$$= \frac{1}{\text{marginal propensity to save}}$$

It will be seen from the above value of the multiplier that it is equal to the inverse of the marginal propensity to save. The lower the marginal propensity to save the higher is the value of the multiplier. Or, in other words, the greater the proportion of income that people spend the greater the multiplier. To secure a rapid increase of income and employment the government should see to it that a large part of invested money goes into the pockets of those who save less and spend more, i.e., employment should be given at the initial stage to those whose marginal propensity to save is low.

Taking advantage of the multiplier process the government can, thus, by the adoption of fiscal measures, remove unemployment or reduce it as much as the technique of production will allow. As we have already said, the value of the multiplier depends on a variety of factors. In the first place, the marginal propensity to consume or to save has been assumed to be invariant with income. This is not correct: as income increases the propensity to consume decreases. Then we have seen that for the multiplier to have full freedom to work itself out credit has to be elastic. Further, the price level has to rise but should not rise very rapidly, otherwise the demand for goods will not increase. If, and so long as, the rise of prices is the effect of increased demand there is no trouble. The price level may then rise to any extent with impunity. But in certain ways the rise of prices is not entirely the effect of increased demand. This statement must be properly explained as it would otherwise be theoretically wrong. For, the supply not changing independently, the rise of prices must be due to and therefore the *effect* of, increased demand. The point to note is that prices may rise due to a shortage of supply (temporary, in certain cases) or due to scarcity of funds or capital-goods. Further, they may rise in the case of some goods due to speculation and the scarcity of labour itself. Labour may not be as mobile as is needed.

What the value of the multiplier turns out to be will also depend on the distribution of income which may change during the operation of the multiplier process. Wages and other incomes rise during the process of income generation and the distribution of income would depend on the elasticity of supply of the factors of production.

But the most important factor to consider is whether the purchasing power increases, and, if it does, to what extent, when the government invests money in a new project. It is in this connection that we have to study the multiplier effect separately for tax-financed and deficit-financed government undertakings.

Deficit-Finance Multiplier

Since the object of the functional-finance is to increase employment through an increase in the demand for goods, it is necessary to adopt ways of financing government projects that will increase the purchasing power of the people. This depends on increasing the supply of money that can be used to buy goods and raise their prices, thereby stimulating production. It is in this connection that we have to examine the relative claims of different ways of financing government projects. Here we shall take up the case of deficit financing of undertakings, i.e., where the cost of production is met by printing notes. In such a case, the budget makes

no provision for meeting this expenditure by taxing people or borrowing money from them. This is why this method of financing expenditure is called deficit financing—the deficit is covered by the issue of notes.

We shall explain the multiplier process with the help of a diagram. The diagram we shall give has no novelty today as it has become as common as the simple supply and demand diagram illustrating how the price of a commodity is determined. We shall have in the diagram a savings curve (in reality a straight line) and an investment curve (also a straight line) set against various levels of income. The investment curve is not shown separately but is shown combined with consumption. For equilibrium, we have to equate, as everywhere else, supply and demand. This is done in this diagram by equating spending plus investment to total income.

Diagrammatical Explanation

We represent income on the X-axis and saving and investment on the Y-axis. C is the consumption curve showing how much is spent out of income. The $C + I$ curve shows how much is spent and invested out

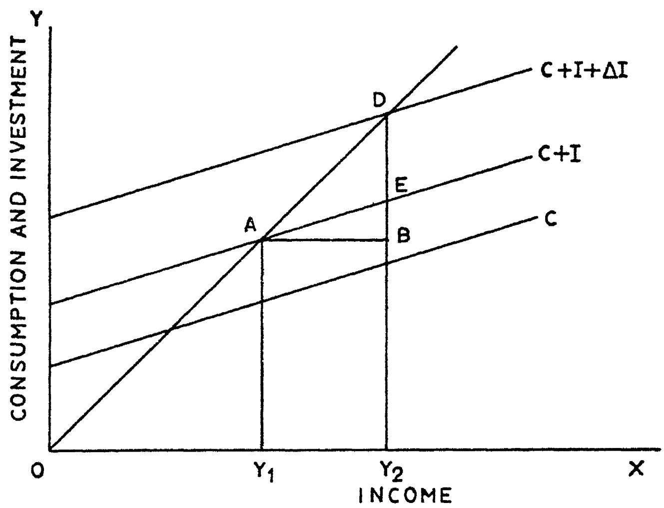


DIAGRAM 16.1

of income. The line OD is drawn at an angle of 45° to the axes. Equilibrium is reached when the entire income generated by the productive system is ploughed back into it. And it comes to be ploughed back when it is either spent on consumption goods (called here consumption) or on capital-goods (called here investment). When income is OY_1, consumption plus investment, as shown by the line $C + I$, is AY_1, which is equal

to OY_1 by virtue of the fact that the line OD makes an angle of 45° with the axes. This means that the system attains equilibrium when the income is OY_1, as the whole income is ploughed back into the system.

When the national income is OY_1 there may be unemployment of some resources and let us suppose that there is underemployment with this equilibrium level of income. Now functional finance is resorted to to increase employment. Let the government print notes and use them to produce some goods. This then is government investment — autonomous investment — that adds to the existing supply of money. Purchasing power increases by the amount of government investment which we may call ΔI. The consumption-plus-investment line is then raised to a new height $C + I + \Delta I$. The vertical distance between this line and the earlier line is constant and equals ΔI, the autonomous investment by the government.

Reasoning as before shows OY_2 to be the new equilibrium level of income. As the income is now greater, it means that employment has increased. If there still remains some unemployed resources, government investment can be increased further. The line $C + I + \Delta I$ would in that case rise to a higher level. The equilibrium level of income would be greater than before and so unemployment would be decreased.

Which is the multiplier in the diagram? We see that income has increased by Y_1Y_2, (which is equal to DB by virtue of the line OD making an angle of 45° with the axes) by investing DE amount of money. DB upon DE is then the multiplier. In this particular figure it is about 1·5. Each rupee invested increases income by 1·5 rupees.

For the sake of simplicity we have assumed that the consumption and investment curves are straight lines. This means that the amount consumed is invariant with income in some sense. We have already observed that this is not what actually happens in an economy. Ordinarily, as income increases the amount consumed becomes a smaller proportion of income. A diagram to represent this fact would have a curved $C + I$, with its concavity downwards. It can be seen from the diagram by imagining such a curved line that the increase of income would in that case be smaller. The multiplier would be a smaller number and it would need greater investment on the part of the government to increase employment.

It will be seen from the diagram that the $C + I$ line does not start from the origin but from a point higher up on the vertical axis. This means that even when the current income is zero there is some consumption. This consumption comes from past income saved. People must live and so they must consume and, if necessary, decrease their cash holding for this purpose. There may be some difficulty in understanding our explanation as, in the diagram, we have not shown the consump-

tion line separately from the investment line. However, for the purpose of understanding the operation of the multiplier effect it makes no difference whether the two are separately shown or not.

Balanced-Budget Multiplier

What happens when government finances its projects keeping the budget balanced has now to be seen. If the budget has to be kept balanced additional taxes must be levied to finance the new projects. In that case, it would appear, employment or income would not increase since no addition is made to purchasing power, the supply of money. That, however, is not quite correct. For, the government-spending of tax revenue becomes at once the disposable income of the people. Assuming that, in the first instance, it goes into the pockets of labourers or those who would spend their entire earnings, the purchasing power of the people is increased by the amount of government investment while taxes levied to get the money do not diminish the purchasing power to that extent. The reason for this is that people do not decrease their expenditure (consumption) to the full extent of the amount they pay in taxes. On the whole, therefore, government investment increases the consumption or the demand for consumption of the people.

What would then be the multiplier can be seen from the following equations. On the Y-axis we measure consumption and investment as before. On the X-axis we measure income. We accordingly call consumption y, income x, investment I, government investment I_g, the proportion of income consumed (the slope of consumption curve) m, and the consumption when income is zero (the intercept of the consumption line on the Y-axis) k. We get the following equations of a consumption curve (a straight line) and the line making an angle of 45° with the axes:

$$y = m\,(x - I_g) + k,$$

and $x = y + I_g + I$ is the equation showing that total income is composed of consumption plus investment.

It will be seen from the equation of the consumption curve that, due to taxation (equal in amount to I_g), the consumption curve falls (x is reduced to $x - I_g$) as the disposable income gets reduced.

The two curves, the consumption curve and the line drawn at an angle of 45°, cut a point where $x = \dfrac{k + I + I_g\,(1 - m)}{1 - m}$

When the people are not taxed to the extent of I_g, and consequently no investment is made by the government, the income x is equal to

$$\frac{k + I}{(1 - m)}$$

Comparing this value of x with its above value we find that there is an increase of income to the extent I_g which is exactly equal to the investment made by the government. The multiplier turns out to be equal to 1. This is the balanced-budget multiplier.

We have already observed that the income multiplier and the employment multiplier are different and will be equal in value only when income and employment increase in the same proportion. This would happen when coefficients of production are such as to make income increase in the proportion in which employment increases. In simple language we can say that when there are constant returns to employment the two multipliers are the same.

Whether the two multipliers will turn out to be the same depends on the technique of production and the proportion of labour to capital already in employment. One thing is, however, obvious to us, namely, that it is always possible to decrease unemployment in times of depression (that is caused by deficient demand) by pumping purchasing power into the economic system. It is more successfully done by deficit financing than by the levy of taxes. It is to be noted that this method of removing unemployment is successful only when unemployment is due to insufficient spending, i.e., when there is excess of available productive capacity, and not when it is due to insufficiency of capital.

Deficiency of Capital

In some of the fairly developed economies the growth of income so changed the pattern of distribution as to reduce the nations' propensity to consume. Incomes went on increasing due to the availability of pure capital and organisation (improved methods of production). But this increase was not matched by the increase of consumption. Such an imbalance always results, or is likely to result, when those who take the decision to produce are not the same people as those who take the decision to consume. In a macroeconomy that is not truly a macroeconomy such things always happen. Hence, in economies that had attained a high degree of development, producers found that they could not sell all that they had produced. The balance between saving and investment had got upset, savings being in excess of investment. The *formal* equality between them was, however, secured by the unsold stock of goods acting as investment. What is not sold or consumed is technically capital and, therefore, a part of investment. It becomes investment of a type that does not contribute to further production of goods. Thereby equilibrium is sought to be maintained as we have investment that will not produce what cannot be sold due to insufficient consumption.

One cause of unemployment is, therefore, deficiency of consumption

or excess of capital (investment) or productive capacity. There can, however, be one other cause of unemployment. Labourers will also remain unemployed when there is dearth of co-operating factors of production. In backward economies this cause is often found to account for unemployment of labour. When income earners spend more and save less than the appropriate amount there results a deficiency of capital. We express this by saying that there is insufficient capital formation. In such cases the economy is unable to produce wealth (income or goods) in sufficient quantity to absorb all available labourers in the productive system. In the case of excess capital, the economy is unable to sell sufficient quantities and, therefore, the available supply of labourers remains unemployed. In one case there is not enough capital, in the other, there is not enough capital in use.

To increase employment in backward economies that suffer from insufficient capital what is needed is to increase capital formation by decreasing consumption and increasing saving. In the beginning of the process of development it is always necessary to save more than what would be natural for people to save. At later stages of development it becomes necessary to consume or spend more than what would be natural for people to spend.

One cannot force anybody directly to save more and it even becomes almost impossible to persuade those who are very poor to save. It is precisely because people are poor that saving is insufficient and it is precisely because they are poor that it is necessary to save more. In other words, to remove poverty we must become temporarily poorer still. However, there is a limit to which we can squeeze money from those who are living on the margin of substinence.

There are two ways of extricating ourselves from the vicious circle in which we are caught when we are poor. First, we must compel the relatively rich people to save more. By means of judiciously devised tax measures we should compel the better-off people to bring down their standard of living. There is, at times, conspicuous consumption or wasteful consumption among the rich. By means of an expenditure-tax or by the taxation of commodities that are consumed by the rich we can, to some extent, increase saving. But much cannot be expected from such a measure. We can, however, get something to invest in the beginning and, if the investment is wise and efficient, it would increase the income of the people and thereby swell the funds out of which savings are made.

The other method is to ask foreigners to save on our behalf, i.e., to borrow money from foreigners and then pay them back gradually in the years to come as and when our income increases. This amounts, in reality, to spreading the sacrifice needed (to save for investment purposes) over a number of years.

Both these methods are generally adopted by backward countries to help them to increase employment. Over and above these ways there are some fiscal measures that can to some extent stimulate saving. One way has already been mentioned by us, namely, taxation of the income or the expenditure of those who are comparatively rich. The other way to persuade them to save is to float high-interest-bearing loans. It is true that much cannot be expected from this method either. For, as economists know today, the rate of interest has very little influence on saving. Not that it has no influence at all; it does make those who are doubting whether to save or to spend their income economise and lend a small amount to the government. But beyond this there is nothing that a little higher rate of interest can do. When government loans are subscribed to, what generally happens is that savings are transferred from private investment to government loans or even from one government investment to another. However, we do not want to belittle whatever importance this fiscal measure might have as a means of encouraging people to spend less and save more. All that we want to say is that when a country is not sufficiently developed it is extremely difficult for it to rely on its own resources for capital formation. A poor country has to depend on foreign loans for its development and for the solution of its problem of unemployment. If foreign loans are difficult to get or if it is thought undesirable to borrow in foreign markets, a poor country must reconcile itself to a painfully slow process of growth and a long period of underemployment of its resources.

Inequality of Wealth and Functional Finance

Howsoever much an individual may like to have more than others, at least a government should not, it is believed, tolerate great inequalities. We shall not discuss the question of the desirability or otherwise of reducing inequalities of wealth and income. But it may be mentioned in passing that when wealth is very unequally distributed it revolts against our sense of justice and equity. A small degree of inequality can be tolerated, nay, it might even be thought desirable. For, the wants of the people and their capacity to derive enjoyment from the satisfaction of their wants are never quite equal. Justice might, therefore, demand unequal distribution of wealth and income rather than equal distribution. There is justice in equitable distribution and not in equal distribution.

However, there is no doubt that in some economies there is much room for bringing about greater equality of wealth than now exists there. Let us see how fiscal measures can help us to effect a more equal distribution of wealth. The obvious fact with which we may start is that,

to reduce inequality, we must either transfer wealth from the rich to the poor or create circumstances in which the earnings of the poor increase with or without a decrease of the earnings of the rich. The latter way of effecting transfer of wealth from the rich to the poor falls within the scope of legislative measures. Minimum wages may be fixed or legislation may be enacted for the grant of other amenities to labourers. Such measures would have the effect, in the long run (other things being the same), of cutting into the incomes of the relatively rich people.

The more direct way of transferring wealth from the rich to the poor lends itself easily to fiscal manipulation. Taxes can be levied on the rich and the proceeds utilised for the benefit of the poor. The needy poor may be given doles or the tax revenue may be used to provide goods and services to them. Incidentally, such a transfer of wealth would also increase the income of the people through the operation of the multiplier. Every wise transfer of wealth to the poor effects a desirable shift of purchasing power from those whose propensity to consume is less to those whose propensity is high. In economies that are stagnating due to deficient effective demand such a transfer, therefore, serves a double purpose: it corrects mal-distribution of wealth and leads to greater employment of labour and, through this, to an increase of national wealth.

CHAPTER 17

Fluctuation and Growth Model Building

What Is It that Fluctuates and Grows?

THIS IS a simple question and might even appear foolish. Yet for an economist it is not a foolish question. When an economist talks of fluctuation or oscillation and growth he has in mind income, employment and other variables of an economic system. But some of these variables are interdependent: income depends upon employment (other things remaining the same) and employment depends upon income. In a similar way, saving depends on income and income depends on saving. Investment depends on income and income depends on investment. In this way most of the variables of our economic system depend on one another. In fact if there is a *system* in the real sense all its components must be interdependent and this interdependence is easy to see when the system is purely endogenous, i.e., when it is free from external influences.

Can we, in such a system, point out any particular thing that varies first and thereby causes variations in other things? If we cannot, the analysis of the problem of fluctuation or growth becomes difficult. But there is a way out of this difficulty. We can begin anywhere we like and assume one particular variable of the system to change first and then see how such a change affects other parts of the system. Moreover, it is possible to consider one change as logically the cause of other changes. For instance, we can say that income flows out of capital and, therefore, logically, capital must change first and income must follow suit. Of course, in all such matters one has to assume that other things remain the same.

Other things never remain the same but for purposes of analysis we have to assume this. We have to separate intermixed effects of simultaneously acting causes.

Thus, in the study of fluctuations and growth we can pick out, say, investment as the first thing to vary. When investment increases (other things remaining the same), income and employment also increase. But the question that can be posed here is: Why does investment increase? Here economists make a distinction between induced and autonomous investment. Autonomous investment can increase without needing any inducement (!). It is that investment, at any rate, that is not induced by a previous increase or expected increase of income. But, one might ask here: Is there any investment that is in every sense of the term autonomous? Autonomous investment, if it is really autonomous, must be self-created, having no cause. Perhaps that revolts against the principle of causation. Hence, by autonomous investment we should understand that investment which is not caused by any change, actual or anticipated, in income. Induced investment is the result of income-variation while autonomous investment is the cause of income-variation.

What is then the answer to our question asked above? What is it that fluctuates or grows? As explained above, all the interdependent parts of our economic system fluctuate and grow. But we economists have singled out one particular variable in terms of the fluctuations or growth of which changes elsewhere are studied. That variable is income — national income, real in the final analysis, though money income can also be taken account of.

Meaning of Fluctuation and Growth

To start with, one might ask why we should talk of growth and ignore decline. This question can be answered by saying that growth inverted is decline and decline inverted is growth. What is a pimple asks one, and the humorous answer given is that it is a dimple turned upside down. That being so, it is sufficient to talk in terms of growth — decline will take care of itself.

The characteristics of growth (or decline) are well known to all though they may not be able to express them in logical language. The essential point is that, in the case of growth (decline), there is a continuous rise (fall) without a reversal. This statement needs elaboration. Growth can be continual instead of continuous but there is a certain element in the case of growth that is continuous. That is the element of stability. It may sound paradoxical to associate growth with stability, with something that remains unchanged. The fact is that (let us talk in terms of income) if there is growth of income there should be growth without decline.

If income goes up and down but shows a steady growth-trend the growth phenomenon refers to the trend of income variation. The trend in such a case continues without a reversal. Wherever there is growth there must be stability appropriate to the sense in which the word growth is used. There must be simple increase without decrease or a simple interrupted increase without a decrease or an increase with temporary setbacks (decreases) with a stable, i.e., continuous, upward trend.

By contrast, in the case of fluctuation there is no such stability. Income increases and decreases or the trend itself shows a rise and a fall. We can express ourselves, then, by saying that in the world of our experience there are continual changes in all phenomena. Income is never stationary; it goes on changing, rising now and falling the next moment. If we fix our attention on the rise-and-fall side or feature of this drama we see fluctuations; but if we fix our attention on the overall-rise or overall-fall side of the same drama we see growth or decline. It is more a question of what we choose to see rather than of what is seeable.

What Causes Fluctuations and Growth?

It is believed that every event, every phenomenon, has a cause. Shorn of all technicalities this statement means that an event is the outcome of a previous event. If *A* is followed by *B* on one occasion it will always be so followed though this sequence may not be evident due to what is known as multiplicity of causes and intermixture of effects. Nothing can happen by itself; an event is, as it were, born to another event. On the principle of causation rests the entire structure of all our sciences.

If income fluctuates or if it grows there must be some factor or factors that can be found responsible for this. Let us concentrate on the growth phenomenon. Suppose income is increasing. Let us ask ourselves why it is increasing. Naturally, if income is increasing the cause of income, that which produces income, must be increasing or expanding. What is it, then, that produces income? As economists we would say that income is produced by the factors of production. If more factors of production are employed income will increase. The word *more* includes the word *better* also. Instead of a larger number of machines, for instance, we may employ a more efficient type of machine. Or, we may say, that other things being the same, the larger the quantities of factors used the greater will be the output.

It is often difficult to say, by merely looking at the factors of production, whether we are using more of them or less. This difficulty is due mainly to the fact that we do not always employ the same kind of factors when we want to increase production. Nor do we always employ the factors of production in the same proportion. We, therefore, resort to

another method of measuring the input of factors of production. We measure them in terms of their money value. In simple words, we take into consideration the amount of money invested in production.

We can, therefore, say with some unjustified complacency, that the cause of growth of income is growing investment. So, if income is growing it is because investment is growing. This is a very simplified statement in regard to the cause of growth. For a thorough study of the problem of growth one must amplify this statement and make a distinction between induced investment and autonomous investment. To say that the cause of increasing income is increasing investment is to say that income increases because autonomous investment increases. For, induced investment is the result and not the cause of income. However, there is a chain reaction between income and investment; income induces investment and that in its turn produces income. But more about this later. Let us be satisfied at this stage with the statement that the cause of growth is increasing investment.

Relationship between Fluctuation and Growth

If the cause of growth is increasing investment, the cause of fluctuation must be fluctuating investment. Most of the words that we use in connection with growth are tricky—investment, saving, over-production and under-production are all tricky to some extent. However, to start with, let us causally connect fluctuation and growth to investment.

Suppose, due to some reason that has its origin either within the system or outside it, investment is increasing, giving rise to increasing national income. If the investment keeps on increasing, other things being the same, income would also keep on increasing. There would then be growth without fluctuations. But if due to some reason investment does not steadily increase, if it begins to decrease after some time, income would also eventually start decreasing. And if after decreasing for a while investment again begins to increase income would again start rising. We would then witness fluctuating income as a result of fluctuating investment.

We have to see whether there is anything within or outside the system due to which investment always increases and decreases by turn. Or is it possible for income to fluctuate even when investment is all the time increasing? It is in this connection that we have to be cautious in the use of the word investment. For, it might at once be stated here that, for the economist, what is produced and not sold becomes a part of investment. But more about this later.

Investment, in the final analysis, is always in real terms, i.e., in order to produce goods and services we have to employ labourers, organisers,

entrepreneurs and capitalists. Employment of capitalists means the use of capital goods. But though investment takes the form of use of factors of production, immediately it consists in the investment of money. It is with money that things are bought or services hired. And when this money gets into the hands of people it begins to circulate, i.e., pass from one hand to another, and the producers have no direct control over how it is used. For the stability of the system or its equilibrium what is necessary is that the money that producers invest (pay out of employed factors) must come back to them for further production of wealth. The system has to be kept running and, if the money that puts the system into motion does not become available for reinvestment, the second round of production cannot begin.

But our socio-economic system is very complicated and at no stage in the movement of money from the producers to the consumers and back to the producers does the economic system have a direct control over it. Money may circulate or it may not; the whole of it may be spent or only a part of it. Due to this liberty that the owner of money enjoys, the decisions of individuals do not necessarily conform to the needs of the whole productive system. It is due to this that investment does not proceed smoothly and the phenomenon of increasing and decreasing investment is witnessed.

So let us say that fluctuations of income are due to the abortive attempt of income to grow and keep on growing. Income grows, because it can only grow, with ups and downs. The features of our economic system that, under the working of economic principles, lead to growth also cause fluctuations. And, as we have observed above, one of the features that is very relevant here is that there are lags and leakages in the movement of money from the producers to the consumers and back to the producers. If we have a model of the economic system it would show why income grows and why it fluctuates. Fluctuations are the short-period phases, as it were, of the phenomenon of growth. This is the relationship of fluctuations to growth.

Limits to Growth

Before we proceed to say something about growth models let us ask ourselves whether there are any limits to growth. Since we have been speaking of growth of income this question must naturally pose itself at this stage. If income comes from the use of productive resources it should logically follow that if there are any limits to growth of income they must relate to the use or supply of resources.

The growth of real income, as against money income, must depend ultimately on real resources. Money, by intervening, helps the system to

make the best possible or worst possible use of resources. If real income is growing it can continue to grow only if real resources continue to increase or if increasing use of resources is continually made. Once resources are fully used, income cannot increase further. This is why we speak of a full-employment ceiling while explaining the upper turning-point.

But over the centuries our real income *has* gone on increasing though there have been lapses indeed. The cause of this must be sought in those changes that have led to a greater supply or a greater use of productive resources. For one thing, population keeps on increasing in most countries. If a larger number of people means a larger supply of labour force and a larger supply of organising ability and, to an extent, a larger number of entrepreneurs willing to bear uncertainty, we have in increasing population an important growth factor. For, in that case all the factors of production would increase and make it possible to increase income. But it is not necessary, when that happens, that per capita income should also increase. For per capita income also to increase production should become more efficient. And otherwise too, the great advances that we have made are due more to better ways of producing wealth than to the mere increase of productive resources.

Let us, therefore, concentrate on efficiency of production as that seems to be the major cause of growing income. And we shall then ask if there are limits to the efficiency of a productive system. Efficiency of production depends on organisation; it depends on how we use the other resources (other than organisation) at our disposal. Natural resources and manpower can be used or misused. How we use them, in what way we combine them, determines what we shall get out of them. This is the work of organisers. It is the organising faculty of man that finds out new and better ways of utilising human and natural factors. Scientific discoveries and the application of such discoveries to production of wealth are made by the intelligence of man. When a man thus uses his intelligence he becomes, for an economist, an organiser. Thus, efficiency of production due to which income increases depends ultimately on organisation. The spectacular advances that we are witnessing today in various branches of production are due to man's inventive faculty which is brought into service by a proper organisation and co-ordination of all available resources. We can, therefore, say that growth depends, in the final analysis, on the factor *organisation*. It is this factor that enables other co-operating factors to put their abilities to effective and profitable use.

Is organisation infinitely elastic, capable of growing in size or in its effectiveness without ever encountering a ceiling? The human mind, whatever the stuff it may be made of, is internally unstable and does not appear ever to reach an equilibrium position. It is expanding, as it were,

from within, answering to the call of exogenous factors. As population increases, as the necessity of increased supply of necessities and comforts of life is felt, the human mind faces a challenge and accepts it. The yet-not-fully-developed mind of man gets a stimulus and under its influence undergoes a process of further development. And so the mental faculties of man go on developing from within themselves and there seems to be no final limit to such a process. This enables us to conclude that there is no ultimate limit to the growth of real income though there may be and must be short-period lapses from continuous growth.

What happens then is that, due to the above continual increase of the supply of the factor organisation, income goes on increasing in the long run. But the increase of this factor is, in a comparative sense, sporadic. It does not necessarily increase just as and when required. Hence, we run short of resources at times and real income stops increasing. This causes not only a stoppage of further investment of money but an actual decrease of it. How this happens we shall see later. Here we are concerned with the mere fact that, as income grows, it faces a temporary shortage of real resources. Hence, income grows with ups and downs.

Equilibrium Growth of Income

Equilibrium is generally associated with absence of change. This is true in one but not in all senses. Take the case of the price of a commodity. Before equilibrium is reached, the price goes on fluctuating; when equilibrium is reached, it becomes stable. Variations of price constitute the attempt of the system to attain the position of equilibrium. But what does equilibrium in its widest sense connote? In the above example equilibrium manifested itself and was, therefore, defined by the absence of expansion or contraction of output of the seller concerned. And a seller stops increasing or decreasing his output when he thinks that by so doing he cannot gain anything further. This is the most important point to note.

If output is increasing or decreasing at a certain rate or at varying rates and the producer thinks that by changing that rate or rates or the acceleration of that rate or rates nothing further can be gained he will be in equilibrium and he will let the output change at the rate at which it is changing. Here, then, there is equilibrium and yet output or the price is changing. But of course we must note the important fact that, while price or output may be changing when there is equilibrium, there must still be something that is not changing. That something is the rate of change or the rate of the rate of change. This is why we said that, in one sense, it is correct, but not in every sense, to say that equilibrium is associated with absence of change.

Growth is itself a changing phenomenon or, to be more correct, a change-phenomenon. But in spite of this change there can be equilibrium if some feature of growth is constant, stable or unchanging. The question that can be asked to elucidate the point is: At what rate should growing income grow so that there may be equilibrium? The question suggests an answer to our former question, the answer that there is equilibrium even when income is increasing if the rate of increase or the rate of change of income is constant. There is another way of answering the same puestion. We can say that a growing economy can be said to be in equilibrium when producers do not find it necessary to change their plans. If income is increasing in a certain fashion, the producers allow it to increase in that fashion. In spite of increasing income the productive machinery is in equilibrium because those who are responsible for production are satisfied with the way things are going. There are various ways of expressing the same idea and so we can put it in yet another way and say that the system is in equilibrium when producers do not find it necessary to alter their plans or to do anything to alter the manner in which income is increasing. And this happens when income is found to behave in the manner in which it was expected to behave.

Why are We Fussy about Equilibrium Rate of Growth?

This may sound a silly question but it is not silly. Why should we take the trouble to determine the equilibrium (sometimes called stable) rate of growth of income? The simple answer is that there is no other growth rate that can be *determined*. We determine the price of a commodity, the supply or demand of a commodity, the rate of interest, the rate of exchange, etc. And when we do determine them it is always the equilibrium values of these things. We determine, for example, the equilibrium price, the equilibrium supply or demand, the equilibrium rate of interest and the equilibrium rate of exchange. Without going into technicalities we can simply say that when we determine the price, for instance, we equate the demand and supply of the commodity concerned. When we equate demand and supply we concentrate on or take account of the position of equilibrium. And the price determined is, therefore, equilibrium price. Is there any other way of determining the price of a commodity? We may *find* the price, we may *guess* the price or we may make a *forecast* of the price; but when we *determine* it, it is always the equilibrium price.

Hence, in the case of growing income, we can determine the equilibrium growth. We can theorise about such a growth and say something sensible about it even when we know that the dynamic state of our

economy will not allow income to grow at that rate. There are many economists who have tackled the same problem; they have built their own models of growth and determined the rate at which income must increase if there has to be equilibrium. But they do not all use the same terminology. Harrod, for instance, uses the word *warranted* in place of the word equilibrium. Then, there are many rates of increase of income about which economists have spoken at length.

As Professor Robbins has said, equilibrium is just equilibrium; there is no penumbra of approbation round this word. When a growing economy is, in the above sense, in equilibrium it does not mean that it is in the most desirable state of the economy. We shall talk at greater length about this in what follows.

Various Growth Rates

The equilibrium rate of growth is called the warranted rate of growth by Harrod. It is the rate at which income must increase if the producers have to be in equilibrium, i.e., when they have to continue to stick to their plans. In yet other words, it is that rate at which income should increase if the actual investment has to turn out to be equal to intended investment. Producers invest money in production in the hope that they will be able to sell what they plan to produce. If they are not able to sell all that they produce some unsold stock remains with them which, technically, becomes investment. So the actual investment (ex-post investment) turns out to be greater than the intended investment (ex-ante investment). So, using this terminology, one can say that the warranted rate of growth of income is that rate which would maintain equality between ex-post and the ex-ante investment.

If and when income happens to increase at that rate it will continue to increase at that rate. This is the meaning of equilibrium. But that is all: there the implications of equilibrium of the warranted rate of Harrod exhaust themselves. As we said above, there is no moral attribute attached to it. From the macro point of view we can judge the desirability or otherwise of a given rate of growth of income. The warranted rate is desirable perhaps from the micro point of view. But this may not be so from the point of view of the whole society or the whole economy. In judging what is the best rate of increase of income we have to depend on some kind of value judgment. Anyway, there are certain accepted criteria of desirability. We want, for example, a state of the economy in which there is as full employment as possible. We might want full employment of human beings or full employment of material resources or full employment of both, to the extent to which that is technically possible. Or we might have in mind some other criterion of welfare. In simple words,

the warranted rate of growth of income is a non-moral concept; it is not immoral, it is just neutral as far as the question of morality or desirability from the social point of view is concerned.

If possible, one would like to have income increase at that equilibrium rate at which all the resources are as fully employed as possible. But that would happen only under certain very fortuitous circumstances. We can have certain other criteria also of desirability: we might prefer maximum per capita income to maximum per capital unit income or even a small income that may keep people more contented than a larger one that makes them feel less contented.

In the context of Harrod's growth model and his warranted rate we may speak of the natural rate also. Preferring one way of defining it, let us say that it is that rate of growth of income that keeps nature most satisfied. That of course is a philosophical way of expressing the simple idea that it is that rate at which natural and human forces would bring about the fullest possible employment of existing resources of all kinds. And it is the natural rate in the sense that there is perhaps a *natural* tendency for the economy to grow at that rate, which tendency is however interfered with by other human tendencies in normal times. As a compromise, we get what we might call the actual rate of growth. This rate is a haphazard rate and we cannot determine it; we cannot have any model that would help us to say what that rate would be because it is an unmodelled rate. By contrast, there is, at least in theory, something definite about the natural rate which has been mentioned above. Harrod calls it the welfare optimum rate as it appears to maximise or optimise social welfare, whatever that might be. If and when the natural rate of growth of income is attained, the resources of all kinds will be fully or optimally used. And in that case it is our mental habit to consider that social welfare is maximum. Professor Joan Robinson calls it the maximum feasible rate and rightly too. For, the best that we can have is when all the resources are fully utilised. It is hard, however, to conceive of an economic order in which the technique of production and the system of remuneration are such as to enable us to utilize, to the full, all the resources simultaneously. But then the best is always difficult to achieve.

The most tractable rate, as we have observed above, is the equilibrium rate of growth. Harrod calls it the warranted rate while Kurihara and some others, like us, call it the equilibrium rate. Professor Joan Robinson calls it the desired rate. It is the desired rate from the point of view of the captains of industry.

Growth Rates, Full Employment and Full Utilisation of Capital

Since it is difficult for all the factors of production to be simultaneously

employed to the fullest extent it is useful to consider the full employment of the factors of production separately. And it has become customary to consider the cases of labour and capital only. There are other factors of production also but capital and labour are, at least objectively, the most prominent. Labour is taken, rightly or wrongly, to include organisation and capital to include the rest. When the task in hand is very complex we have to make simplifying assumptions. The danger involved in such simplifications is minimised when we remember that our conclusions are coloured by and dependent on these simplifications.

Harrod's warranted rate of growth implies full employment of capital; it implies that the investment made in production does not lie idle. Why is this the implication in Harrod's warranted rate of growth of income? Harrod's warranted rate is, as we have explained above, the equilibrium rate, i.e., it is the râte at which income should increase if the producers have to be in equilibrium. And they are in equilibrium when they find that they have been able to use their capital as fully as they had wanted to. Their investment should not remain unused. Nor should it be over-used. It might appear that investment cannot be over-used, but when the inventory shrinks below the planned level we say that we have over-used our investment. However, we are concerned with the simple fact that producers are in equilibrium when they find that their capital-stock is fully utilised. It is only then that they will not alter their plans and, if they do not, the economy proceeds in the manner desired by them. There would then be no unpredictable element in the situation. There is equilibrium in the sense in which the word is used in this context.

We should not, however, misunderstand what has been said above. Harrod's warranted rate does not imply that as much capital is used as is socially, or in some other sense, the most desirable. It only implies that whatever amount is invested by the producers, whatever capital is possessed by the producers, is actively used for purposes of production. We do not have to witness the phenomenon of machines and other equipment lying idle in factories. That is all the implication of full employment of capital.

In Domar's model, full employment of labour is assumed and Domar goes on to determine conditions in which full labour employment would be maintained. But one has always to face a difficult task when one assumes full employment of labour because population has a natural tendency to increase in most economies. Capital too, in a way, has a tendency to increase and it has very much increased but there is an important difference between the increase of capital and the increase of population. Capital is man-made and its increase is under the control of man. It increases, we might say, endogenously and its growth is, therefore, according to the wish of the economic system; it increases to the extent to which the productive system wants it to increase. This

statement has to be modified to a certain extent. There are factors over which an economy has very little control. Decisions about production are taken by individual producers and these individual decisions may and often do come into conflict. Due to this lack of co-ordination individual producers find that they have happened to create a larger or sometimes a smaller amount of capital than they wanted to. Apart from this, however, the supply of capital is within human control and there is no exogenous factor that determines it.

The supply of labour is not under the control of an economic system to the same extent. Though population is also, in a way, a man-made factor, it is not the outcome of the rational mind of man. It grows in a fashion that is not determined by the needs of the productive system. The growth of population is, therefore, treated as an exogenous change. This difference between capital and labour makes the work of those model builders difficult who proceed on the basis of full employment of labour.

While we are at it, let us note the fact that the rates of growth of the two factors, labour and capital, do not harmonise, for the reason explained above. It may be that capital accumulation outruns population or population outruns capital accumulation. What actually happens or is most likely to happen will depend upon the determinants of capital accumulation. Does it depend at least partly on the savings of the people or on the productivity of capital? Is productivity determined for the purpose in hand by the share of profit in total income? These are the questions that need to be answered by the model builders.

Model Building

Model building is not a new enterpirse. The physiocrats were perhaps the first to build a good macro-model of an economic system. Quesnay's *Tableau Economique* shows how wealth circulates and if this circulation proceeds without disturbance we would have a stable economy. The physiocrats explained the working of the economic system by taking the simple case of two classes, one the proprietor class and the other the cultivator class. All the wealth or money that is pumped into circulation must keep on circulating—from the proprietors to the farmers and from them ultimately back to the proprietors. The *tableau* is a simple macro-model and our modern models are basically the same as that model. What is required for equilibrium is that money that lubricates the wheels of the economic machinery should keep on circulating. Its velocity should not decrease, at no stage should money or purchasing power become sterile. Harrod's model and Domar's model and all other models have to depend on this central requirement, namely, that if there is to be equilibrium (whether income is increasing or not) there should be no

sticking of money or purchasing power at any point in its circular flow.

However that may be, we are here concerned with the art of model building. We can have, if we like, a model of an individual economic unit but there is no novelty in this. When we say that an individual maximises his net income or net utility and explain how he does it we have a micro-model. Such a model is not difficult to understand nor difficult to construct. But our concern has been with the working of the whole economic system. The economy of a country is complex, it is composed of the innumerable economies of small units. We have insufficient knowledge of how those individual economies combine and how the macro-economy works and what causes disturbances to its smooth working and what ensures or would ensure its stable progress. Hence, our models are macro-models. We build models of the whole economy and thus face up to a gigantic task. How can we have a model of such a complicated machine? Certainly the task is difficult. But we simplify it by making just a model of it. A model should depict all the essential features of an economy—we can let the details go, the details which do not matter for us or which are assumed not to matter. We have the liberty to decide which features of the macroeconomy are essential and which are only superficial. And the purpose of model building helps us in picking out the essential features. The main purpose for which we build models is to find out the cause or causes of the observed fluctuations of income or its growth. Economists had observed the cyclical movement of economic activities and the see-saw movement of national income. They wanted to know what precisely was the cause of this and whether there was anything within an economic system that generated such cyclical changes. Then came the growth economists who, knowing that national income was on the increase, wanted to gain further knowledge about the forces that caused growth and the manner in which growth should proceed if it was to be stable. We shall explain later on what precisely stability of growth means. Just as we determine the equilibrium price of a commodity or the equilibrium supply of a commodity, similarly we determine the equilibrium (increasing) income of an economy. It is with these objects that economists constructed models and for that purpose selected some important and essential variables to serve as raw materials.

All our modern models take their cue from the new economics of J. M. Keynes. He can be regarded as the first economist who dealt with the problems of macroeconomy in a systematic way. Savings, investment, employment, income, consumption and interest rates figure prominently in his economics. The model builders of today have selected their variables from among them.

In the simplest models, income is taken to be the only variable. Since we are concerned with the case of increasing income we must take the incomes of at least two periods into account. The amount by which

income has increased from the first to the second period has, therefore, to be known. The initial increase of income is thus a given datum. Then, there are other known factors and in the simple models they relate to the behaviour of producers (investors) and the behaviour of consumers (or savers). Their behaviour is expressed as their reaction to their incomes. In simple words, the behaviour of a producer is indicated by the investment function, a function that shows the relationship between the increase of income (of the current period over that of the past or of the past period over that of the period prior to that) and new investment. The behaviour of consumers is indicated by the consumption or the saving function, a function that shows the relationship between income and consumption or income and saving.

These behaviour equations are very important. The equilibrium growth of income given by solving these equations depends on what precisely the structure of these equations is. But what must be noted here is the fact that the behaviour of investors or of consumers (savers) is shown as depending on their income or change of income. As a matter of fact, there are so many other things on which investment and consumption or saving depends. But the tacit assumption here is either that investment and consumption depend only on income, total or marginal (or if they depend on other factors their influence is negligible), or that the other changes on which their behaviour depends need not be separately taken account of as they exercise their influence through their effect on income.

Hence, in very simple models, we begin with the income of two or more given initial periods and the known reaction of investors and savers (they are also consumers) to income, or changes in income. We can then begin to solve the equations to get the equilibrium growth of income. This rate of growth is the one that, if and when attained, would keep the investors satisfied with their plans. Or, in other words, at this rate of growth of income the ex-post investment would be equal to the ex-ante investment. How this is the implication we will presently see.

The equation for consumers' or savers' behaviour shows the amount that would be saved out of income. The investors' behaviour shows how much they would invest. In equilibrium, the investors find their ex-post investment equal to ex-ante investment. But ex-post investment is the same thing as ex-post saving. And, by hypothesis, ex-post saving is equal to ex-ante saving. Hence, what is required for equilibrium is that the saving shown by the equation of savers' behaviour should be equal to the investment shown by the equation of investors' behaviour. This is how we proceed to determine the equilibrium rate of growth of income. And, as we have already said, it is only the equilibrium rate that can be determined and it is that rate which is all the time trying to assert itself. The actual rate, which is the result of so many causes acting on the system

at every moment of time, shows deviations from the equilibrium rate—it annoys, as it were, the equilibrium rate and makes it assert itself. In simple words, producers want to be in equilibrium and so they will do everything possible to bring about the equilibrium rate of increase of income.

Let us very briefly illustrate this point with the help of the simple model of Harrod, the pioneer in growth models. His two equations showing the behaviour of consumers (savers) and producers (investors) are:

$$S_t = sY_t \tag{1}$$

$$I_t = g\ [Y_t - Y_{t-1}] \tag{2}$$

Equation (1) is the savings function, showing that in period t, income earners save S_t which is an s fraction of their income in the period t. Two facts have to be noted: first, the savings of period t depend on the income of period t; second, the proportion of income saved is constant. This latter, more than the first, is unrealistic. As income increases we save a larger portion of it. Therefore, s is, an increasing function of income, Y. But it is taken to be constant to start with and this simplifying assumption is relaxed later on. But for the purpose of determining the equilibrium rate of growth it does not matter. The broad features of the economy are not altered by this assumption, nor is the conclusion qualitatively altered by it.

There is yet another fact that may be noted. In the savings equation, S_t stands for actual as well as intended saving. This means that it is assumed that income earners stick to their decision about saving: if income earned does not come up to their expectations, they spend less but do not cut down their savings. This is not a very realistic assumption. But an assumption is only an assumption and it is always made with a purpose. What really happens is that when income does not come up to their expectations they spend a little less and save a little less. There may be some people who would throw the whole burden on their savings. But, as it is not known what precisely is the attitude of consumers, one has to make some convenient assumption. Econometricians can tackle this problem. But for the purpose of finding some general behaviour of growing income that would keep the system in equilibrium it does not much matter if our assumption in regard to saving is not very realistic. As we shall see, some other models also yield the same condition for equilibrium growth as Harrod's. This proves that the assumption in regard to saving does not materially alter the conclusion.

Then we have the investment function in which I_t, the intended investment in period t, is equal to a multiple g of the difference between the income of the period t and the income of the previous period. In other words, the producers' intention is to invest g times the increment of income. The significance of the word intention rests on the significance

of the word investment. If investment were to mean only what the producers spend on production at the beginning of period t or during that period, then there would be no sense in speaking of intended investment. For, the intended and the actual would not then be different. But, as we have observed earlier, investment includes the goods that were produced to be sold but could not be sold. That is a part of investment for the reason that there is no other head under which it can be put. Goods not sold are, in a way, saved. And since producers as producers are not consumers their saving is technically called investment.

Hence, I_t is the intended investment which must equal actual investment if there has to be equilibrium. We, therefore, put it as equal to S_t which is actual (also intended) saving where actual saving is identical with actual investment.

We, therefore, put savings as equal to investment from our equations and get the condition for equilibrium growth of income. It turns out to be

$$\frac{Y_t - Y_{t-1}}{Y_t} = \frac{s}{g}$$

The rate at which income should increase to keep producers satisfied and, therefore, to maintain equilibrium in the sense in which that word has been used by us is given by s/g. Let us remember that s is the saving ratio. It is known as the multiplier factor, g is the investment coefficient. It is known as the accelerator factor. We shall have more about these words in the chapters on the Multiplier and the Accelerator and so we do not need to say more about them here.

When the saving and the investment coefficients s and g are taken to be constant, the rate of increase of income in equilibrium is also constant. When s increases with the increase of income, this rate tends to increase. When g increases with the increase of income, this rate tends to decrease. If with growing income productivity of capital (marginal efficiency of capital) decreases, the producers would have to increase their investment to produce a given amount of income. Hence, the value of g would tend to increase with the increase of income. The value of s/g would then depend upon the extent to which these coefficients change. But one fact to be taken note of is that the rate of equilibrium growth of income is positive. This means that for equilibrium purposes income should keep on growing. In other words, growth can be sustained by growth only. If there is any disturbance, this growth will be retarded or accelerated and then fluctuations in income would be witnessed. Harrod explains this phenomenon also but we are not concerned with that here. In the world in which we are living things never proceed smoothly. Our systems are never fully endogenous; they are subject to exogenous forces. Our systems are never fully closed and, therefore, there are always some disturbing factors in the situation. Equilibrium, as we have observed

earlier, is therefore hard to achieve. If income is increasing it must go on increasing if there is to be equilibrium, if there is to be stability. The rate at which it should continue to increase is dependent on the precise form and structure of the saving function and the investment function. But the broad fact is that income must increase, growth must continue if it has to be stable in some sense.

The model described above is a very simple one that we have chosen to show how model building is done. There are many assumptions in this model and, though they are not very material for our purpose, it is interesting to know that we can make some changes in these assumptions and get models of a somewhat different type. Much of the work in model building is done for its own sake—it is enjoyable, it is directly enjoyable as consumption is. We must remember that a model of growth should also be capable of showing how and for what reasons income fluctuates; our models should be models of growth and models of the trade cycle. We have already explained how the two changes go on simultaneously. Income grows but unevenly, with ups and downs. And if our models can show that, we are satisfied. But the reality is too complex to allow us to build a model that would exhibit all the features of our economy. We do not want such a model and we do not hope to be able to construct such a model.

How We Change Our Models

We can change our models by changing our equations. In the two behaviour equations we have made saving and investment some kind of functions of income or increment of income. In Harrod's model considered above, saving is a function of current income and investment a function of current increment of income. Income or its variation is perhaps the most important single factor on which the amount saved and the amount invested depend. We have said *perhaps* because in some cases some other factors may for a while exercise a greater influence than income or its variation. The other factors that influence the behaviour of consumers are the level of prices expected to rule in the future, the amount of saving already made in the past and the expected change in the standard of living. Investment, likewise, depends on the capital that has already been built up and its rising or falling productivity, the rate of interest, the prices of other factors of production, imports or changes in imports and government spending. We can incorporate some of these variables into our system and accordingly change our equations. We can also change our models without including these variables by simply changing the periods of time on which saving or investment is taken to depend. We can also replace the saving function by the consumption function, and

make consumption in one period a given proportion of the income of the previous period. In such a behaviour-equation, consumption may be taken to be fixed in the sense that ex-post consumption is the same as ex-ante consumption. Harrod assumed that income-earners do not change their decisions in regard to saving; here we can assume that they do not change their decisions in regard to consumption.

The results we get by making the above changes do not differ in their broad features from what we get in the case of Harrod's model. But changes in premises must have their counterpart in changes in conclusions. But we are satisfied if we get the result that makes stability of growth dependent on continued growth of income.

So there is no end to the variety of models that one can build. It is interesting, however, to incorporate the influence of accumulation of savings and the accumulation of capital—the first, in the consumption function, and the second, in the investment function. As one finds that one has more savings than before one feels less inclined to save a further amount. Likewise, as one finds that one has accumulated much capital one feels less inclined to accumulate more. This may be with the object of minimising risk or because of the diminishing marginal efficiency of capital.

We can incorporate the influence of capital stock on investment in the following way:

$$I_t = g\,[Y_t - Y_{t-1}] + g'.K_{t-1}$$

K_{t-1} is the stock of capital that is already in existence. The suffix of K (capital) is, therefore, $t-1$. Or as a variant, for the purpose of simplification, we may replace K_{t-1} by K_t.

The sign of g' depends on two considerations. First, on the psychology of producers pure and simple and, second, on the phase of the cycle (a cycle being a short-period phenomenon in the process of growth). To explain these points, producers may like to invest more as investment increases—not because more goods have to be produced as that is covered by $g\,[Y_t - Y_{t-1}]$ but simply because the accumulated stock of capital makes them feel that they should expedite investment—or they may like to invest less, thinking that enough has already been done as far as production goes. This is the pure psychological effect of capital stock on investment.

Then there is the psychological effect via the phase of the cycle. During a depression the stock appears too big and, therefore, it makes the producers less willing to increase investment, while during a boom the stock appears small, making the producers very enthusiastic about investment. The sign of g' would, therefore, depend on the above considerations.

What is capital stock for producers is savings for consumers. While saving out of income many factors are taken into account. It is true

that habit plays an important part in determining the percentage saved, but size of income, expected income in the future, demands on future income and the saving that has already been made, all have an influence on the decisions of savers. It is, therefore, not unrealistic to make S_t equal not only to sY_t but to $sY_t + s'.V_{t-1}$, where V_{t-1} is the accumulated stock of savings. Since what is actually saved becomes realised investment we can substitute $s'.K_{t-1}$ for $s'.V_{t-1}$. The sign of s' would depend on the psychology of savers. It is not unreasonable to believe that it is most likely to be negative.

We can introduce certain further modifications in models, thereby making them, of course, more complicated. For instance, we can make the consumption function and the accelerator non-linear. We can have distributed time lags (e.g., consumption may be made to depend on or be a function of a number of past periods and so also the investment function). We can introduce exogenous factors by incorporating autonomous elements in the consumption and investment functions. We can make saving and investment depend not only on the current and past events but on anticipated future events also. We can open out our system to foreign influences by taking account of imports and exports. All these variations would have the effect of making our models more complicated though, in certain respects, they might become more realistic.

We shall not give these models here. For, our object is to say something about model *building* and not about models themselves. But since we have said, by way of explaining model building, something about Harrod's model, we might as well make a mention of Domar's model. These two models are very similar though their approaches are different. Domar equates supply and demand to get the equilibrium rate of growth of income. Equilibrium always requires a balance between supply and demand. When some money is invested in production it simultaneously (with an unavoidable time gap, of course) creates money income and real income, i.e., demand for goods and a supply of goods. These have to be equal if there has to be equilibrium. When supply is equal to demand, the savings of the people will also be equal to investment and so Harrod's condition for warranted rate of growth is satisfied. But Domar proceeds from the angle of demand and supply.

To use his terminology, let the savings ratio be α and σ the output-capital ratio, ordinarily known as the productivity of capital or investment. Y is income, ΔY is the increment of income, I is investment and increment of investment is ΔI.

As $\frac{1}{\alpha}$ is the multiplier coefficient the investment of ΔI will create an income of $\Delta I/\alpha$. Since σ is the productivity of capital the addition made to supply of goods is equal to σI. Putting demand equal to supply we get

$\dfrac{\Delta I}{\alpha} = \sigma I$ $\left(\Delta Y = \sigma I \text{ since } \dfrac{\Delta Y}{I} = \sigma. \text{ Also } \Delta Y = \dfrac{\Delta I}{\alpha}, \therefore \dfrac{\Delta I}{\alpha} = \sigma I\right)$. Further, since the amount saved (αY) is equal to the amount invested (I), we get $\alpha Y = I$. From the above equalities we get the rate of increase of income in equilibrium to be $\dfrac{\Delta Y}{Y} = \alpha\sigma$.

This equilibrium rate is the same as Harrod's warranted rate which can be seen from the fact that, in Harrod's notations, $\alpha = s$ and $\sigma = 1/g$. The capital-output ratio $\left(\dfrac{1}{\sigma}\right)$ tends to equal, in the long run, g which is the acceleration coefficient in Harrod's model. σ is the output-capital ratio in Domar's model. Thus, $s/g = \alpha\sigma$.

In Harrod's model, increment of income leads to investment. In Domar's model, investment leads to increase of income. The cause-effect relationship is reversed in Domar's model. But in mathematical equations this relationship is not shown. One thing is shown to be associated with or equal to another thing without indicating the time sequence involved. This is because our solutions yield us the end result, the equilibrium position, only.

Growth and Stability

Stability implies absence of change. And, therefore, equilibrium and stability mean fundamentally the same thing. We have had occasion to say that even when some economic entity is changing it can still be associated with the notion of equilibrium when its rate of change is constant or the rate of the rate of change is constant. In the final analysis, something must be constant if there has to be equilibrium. Very much the same thing can be said about stability.

In the literature of economics the word stability has been used in another sense also. A position is said to be stable not only when it does not change but when the balance is such as to regain itself if and when it is disturbed. Take the case of equilibrium of supply and demand. In Diagram 17.1 are represented four cases in which the supply curve cuts or meets the demand curve. In both the graphs the demand curve is sloping downward. In the first, the supply curves cut the demand curve, while in the second the supply curves are tangential to the demand curve. The supply curve S_1 cuts the demand curve at point *a* where the equilibrium is said to be stable because any movement from the point of equilibrium releases forces that tend to re-establish equilibrium. If there is a movement to the right of *a*, cost is in excess of price and so production shrinks till the equilibrium point is reached again. If there is a movement to the left of point *a*, cost is below the price and that

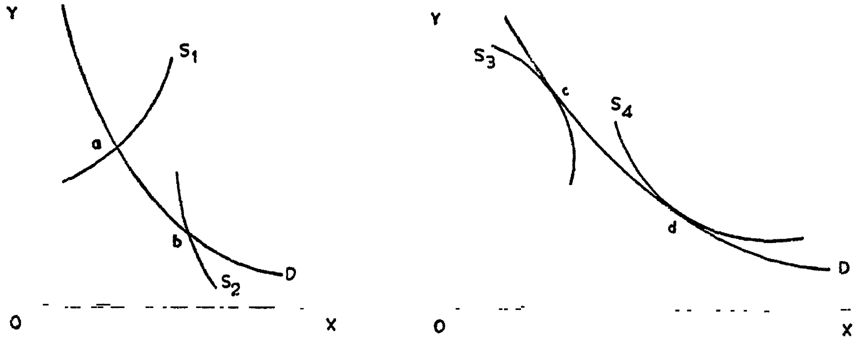


DIAGRAM 17.1

leads to an increase of production till the point of equilibrium is again reached. Hence, the point *a* is said to be a point of stable equilibrium.

Arguments similar to the above would show that the point *b* is one of unstable equilibrium. The point *c* in the second diagram is one of partial stability because from the left of point *c* the movement is towards the equilibrium point while from the right of it the movement is away from equilibrium. Similarly, the point *d* is also one of partial stability.

All the four points indicate positions of equilibrium, i.e., unless an exogenous force comes to be applied the equilibrium maintains itself. The question of its instability arises only when an external force impinges itself on the system. An equilibrium is described as stable when the forces that constitute this equilibrium react to the exogenous stimulus to neutralise its effect. When, however, the exogenous force is able to dominate, as it were, over the harmony of forces that make for equilibrium, the position is said to be unstable.

This reasoning can be applied to the case of income also. Income may be constant or increasing. In the former case, the word stability qualifies absolute income; in the latter, it qualifies relative income or, in the context of the present discussion, the rate of change of income.

Stability indicates either that income remains at a constant level (or, if it changes, it tends to regain its former level) or maintains its rate of growth. When we talk in terms of equilibrium rate of growth, stability refers to the preservation of such a rate.

Income (growing income in the context of present discussion) is unstable when there are fluctuations in it or when the deviations from the point of equilibrium tend to get widened. Harrod explains how there can be instability in this sense.

Whether income will increase at the equilibrium rate or not depends on a number of factors. In the simple model of the type that we have selected for the explanation of how we can determine the equilibrium rate of growth, the growth rate depends on the values of the multiplier

and accelerator coefficients (s and g or α and σ) and the time lags in the adjustment of consumption and investment to changes of income. Certain values of the coefficients would enable the system to attain the equilibrium rate of growth. If we could ensure those values we would enable the system to attain the equilibrium rate of growth. In that case we would say that the system is inherently or endogenously stable. When the stable growth rate is attained, it is only an exogenous force that can upset it. But if that force exhausts itself, the equilibrium position will once again be established as the system is endogenously stable. But whether the exogenous force will exhaust itself or not is the question. If it could release certain psychological reactions it would find sustenance for itself and, consequently, the growth of income would become unstable. Harrod relies upon such psychological reactions of producers for his explanation of the instability of the system.

CHAPTER 18

Stock and Flow Concepts

Relationship of Stock to Flow

IN ECONOMICS as in physical sciences we make use of these concepts. In a very general way we can say that they correspond to matter and energy. What we see and sense is matter; we cannot see energy though we can sense it. It is difficult to define these concepts but that matters little for our purpose. Matter, we might say, is condensed energy. So long as energy flows evenly and continuously it does not turn into matter. When its flow is uneven or interrupted it changes itself to form matter. This is not a scientific explanation of the relationship of one to the other but it is useful for our purpose. Physicists say that matter and energy are convertible. Matter is a storehouse, as it were, of energy. We might also say that matter in its most rarified form is energy and energy in its most condensed form is matter.

What is it That is Stocked and What is it That Flows?

As economists, we can say that what is stocked is a material thing and what flows is a service. From this, by a metamorphosis, it follows that matter is a stock and service a flow. A labourer is a stock, labour is a flow. A capital good is a stock, the service it renders is a flow. A consumption good is a stock, the utility of it that the consumer enjoys is a flow.

Services of factors of production constitute flows; they are, therefore, a form of energy. But the factors themselves are stocks and, therefore, belong to the category, matter. The difference is not so simple when we come to the concept of money for the reason that the word money has been used in more than one sense. Fisher uses it in one sense and Keynes, following the Cambridge tradition, uses it in another sense. But we shall take up this question a little later.

Use of These Concepts in Economics

These concepts have played a part mainly in the theory of interest and in discussion of the fundamental motive of a producer.

Interest can be variously defined but for the purpose in hand it is best to define it as the price of money. In the most general sense, interest is the income of the capitalist, i.e., of one who waits for or postpones consumption. In the type of economy in which we live this waiting is done through the use or surrendered use of money.

Interest, when thus tacked on to the concept of money, has been defined in three principal ways. It is said to be determined by the supply and demand of money that is saved, by the supply and demand of money that is lent, and by the supply and demand of money (that is held). We can discern in these theories two different concepts of money — the flow concept and the stock concept. For Keynes, who says that interest is determined by the supply of money, money is that which is *held* in liquid form. It is a *stock* of wealth that can be used to buy goods and services. For others (Fisher as an example), money is that which is actually used as purchasing power. It is what *flows* out of the stock of money in the possession of one.

For Keynes, therefore, money is a stock whereas for Fisher and some others it is a flow. Money for Keynes has, thus, no velocity. For Fisher it necessarily has velocity. The effective supply of money as a flow is, therefore, determined by the amount of money that is used as purchasing power multiplied by its appropriate velocity of circulation.

Here by flow we do not understand energy or mere service. But the service concept is there all the same. Unless money serves you as purchasing power it is not money. It must have velocity; it must be on the move; it must possess the characteristics of a flow. This difference in the precise connotation of the word money does not make much of a difference in the theory of interest. But the dressing of the theory does depend on what notion one has of money. Otherwise all roads lead to Rome. The various elements of a system are interlinked and the equilibrium rate of interest is reached when the supply of and demand for money, howsoever defined, are equal. The supply of money saved equals

the demand for money saved: the supply of money lent equals the demand for money borrowed: the supply of money held (liquid) equals the demand for money to be held.

Which theory of interest is logically the most correct (logically, not practically) must be judged by its application to the case of Robinson Crusoe. Keynes's theory is most suitable for an explanation of market phenomena. But we are not concerned with interest theories as such. It is sufficient for our purpose to note that money can be held as a stock and that it can be used as an active flow. And in theory the question might pose itself whether money is really money when it is simply held in a liquid state.

Profit and Loss versus Balance Sheet Approach

What does a producer want to do? Does he want to maximise profit (excess of income over cost) or is he concerned with the shaping of his balance sheet in accordance with his asset preference? Is a producer concerned with income or is he concerned with capital? Is he interested in his assets or is he interested in the yield of his assets? These questions bring to light the possibility of alternative approaches to the theory of production. When one is interested in one's assets, we can say one takes the stock view of one's concern. Likewise, when one is interested in one's income one can be said to be taking the flow view of one's concern.

In traditional economics it was taken for granted that a producer's main concern was with profit. But this word was used in the colloquial sense, i.e., as the excess of income over cost. And income meant money income and cost meant money expenditure. Later, as a refinement of this view, it was stressed by some that income should include all that the producer considers desirable and cost should include all that he considers undesirable. This meant the inclusion of psychic income and psychic cost on the one hand and, to an extent, the taking of a long period of time into consideration, on the other. However, it still meant that a producer was aiming at maximum *income*. The flow concept was preserved: a producer was assumed to think only in terms of profit and loss.

Professor K. E. Boulding challenged this view and maintained that a firm had not only a profit and loss account, it had a balance sheet also. A firm was not only interested in income, it was also interested in its capital, its assets. This view involved a criticism of the maximising postulate of traditional economics. Focussing attention on the balance sheet, Boulding said that a firm's concern was with the maintenance of the various assets in certain preferred proportions. In other words, a firm wants not maximum net income but the homeostasis of its

balance sheet. If, for instance, the price of its product rises the firm increases production, not to maximise net income but to bring about or restore the previous ratio between assets held in cash and those held in the form of finished goods. This was an alternative way of explaining the behaviour of a producer. But questions of great theoretical interest arise here. As we are here concerned with the use of stock and flow concepts in economic theory we shall simply ask ourselves a couple of questions.

The main question is this: Does a producer think in terms of income or in terms of capital (assets)? Does he want maximum income (per unit of time!) or does he want maximum capital, wealth or assets? If the former, he is for a flow; if the latter, he is for a stock. The question that comes up here as a corollary is: If a producer wants maximum income is it for income's sake or for the sake of capital that it would help to build up? If, on the other hand, a producer wants maximum capital is it for the sake of capital as such or for the sake of income that would flow from it?

There are two types of people. First, there are those who love to possess, to store and take delight in feeling that they have a command over so much wealth. They may be called the miserly type for want of a better word. Then, there are those who love to spend and get enjoyment by converting purchasing power into a flow of utility. Such people may be called the spendthrift type. It is perhaps not possible to find a pure type of either a miser or a spendthrift. They set the limits within which a person can, according to his temperament, be located. Some of us, thus, would care more for capital than for income while others would care more for income. A theory that can combine the classical view with Boulding's would, therefore, come closer to the facts of life than one that depended on either of these views.

Ability To Pay a Tax: A Stock or a Flow?

Directly from what has been just discussed follows the consideration of the concept of ability to pay a tax. The most fundamental principle of taxation is that of the least aggregate sacrifice or, what might better be called, minimum social sacrifice. In matters of taxation the point of view of an individual must subordinate itself to that of society, whatever that word might connote. A tax or taxation in general must be so devised and imposed as to reduce to the minimum the sacrifice made by the whole society. It is an extremely difficult task to measure social welfare or illfare. There are theoretical as well as practical difficulties inasmuch as one does not know what social welfare ultimately consists in. One must first know what a society is before one can say what its

suffering or its welfare is. Much attention has not been paid to this question and the discussion of social welfare or social sacrifice has proceeded on the tacit assumption that there is no difficulty about the meanings of these words.

Anyway, as far as individual taxes are concerned, it is believed that a tax must be made proportional to the ability of a taxpayer to bear the tax burden. A rich person has greater ability to bear a tax burden than a poor person. But what is the measure of poverty or of affluence? It is here that we come across the concepts of stock and flow.

Most of our direct taxes take account of income on the tacit assumption that income is an index to a person's ability to bear the burden of a tax. As a typical example, we have the income tax. Its rate is made to vary with the size of the total income of the tax-payer. There are different ways in which the rate can be made to vary: it may vary proportionately, it might vary progressively or it may vary regressively. No attempt is made today to make a tax regressive in its incidence, though in its operation it might turn out to be regressive. Everywhere, income tax is levied at a progressive rate, the rate being more or less arbitrarily determined. The simple idea is that a wealthy man suffers less than a poor man in paying the same amount as tax. And the suffering decreases progressively as wealth increases. But here, by the expression a wealthy man we understand one who has a larger income. This implies that the ability to pay a tax is taken to be a flow. For, income is a flow and the greater the size of this flow the greater the ability to pay a tax is supposed to be.

But we have some taxes that are levied on the wealth or assets of a person such as wealth tax or a capital levy. A capital levy is a tax which is made use of in some kind of an emergency. Lately, we have witnessed a regular use of a tax on wealth, such as the wealth tax in our country. There are a number of reasons for the levy of this tax but we are concerned with only one of them. Wealth, it is believed, is also an index of ability. While income is ability in the form of a flow, wealth is ability in the form of a stock. A man who has no income, for example, can yet pay a tax if he has accumulated wealth. A man with greater income may have a smaller ability if he has much less wealth. For one thing, income flows out of wealth (provided we include in wealth and income their psychic contents), for another, wealth could be utilised to pay tax, in which case wealth can be taken to serve the purpose of income.

Where the income of the people is proportional to their wealth it matters little whether taxes are levied on income or wealth. If a tax is small, it is in all normal circumstances paid out of income. If it is a big amount it might encroach on the wealth of the payer. But that would not matter as far as our object is to equalise the sacrifices of the tax payers.

But incomes are not in all cases proportional to wealth. For this reason it becomes necessary to make a distinction between income tax and wealth tax. And, making such a distinction, the two can be judiciously combined to approximate the real burden of taxation as closely to ability as possible. That, however, must remain a difficult task. For, there are some who suffer more when their wealth decreases than others. They belong to the miser-type; for them wealth is not only a source of income, it is a valuable possession in itself.

One more fact may be noted in passing. Wealth or capital is not necessarily in the form of physical, material things. It also exists in the form of a man's personal qualities. It is spoken of as personal capital. A tax-payer may have the same income and wealth as another but he may be stronger and healthier. He has, therefore, a more abundant *source* of income. In paying a certain amount of tax such a man suffers less than another who is weaker and older.

Total and Marginal Utility and Productivity: Stock and Flow Concepts

The utility of a stock of a commodity is total utility. The utility of a small part of that stock is marginal utility. If you begin with one unit of a commodity and increase it to two units the addition made by the second unit to the utility of the first is called the marginal utility of the stock consisting of two units. It will be seen that total utility is thus made up of marginal utilities. All marginal utilities concentrated or condensed into one lump make up total utility. Thus understood, total utility is a stock concept while marginal utility is a flow concept. If we take a durable commodity, for instance an electric fan, the utility we get from its use hour by hour can be called its marginal utility. The utility we get from it day by day can also be regarded as its marginal utility. But the fan as a whole represents so much utility stocked in a material mould. Utility has its *stock* and *flow* — there is a stock concept and a flow concept of utility.

All that has been said about utility applies to productivity also. We have marginal productivity of capital, marginal productivity of investment, and marginal efficiency of capital. All these are flow entities; they refer to a flow of income from a stock of capital. A capital good is a capital-good only by virtue of the fact that it enables one to get a flow of income from it. It is, therefore, a stock of income. The value of a capital good is calculated by adding up the discounted marginal productivities.

One word might be said about the relationship of a stock and a flow to the time element. A flow has its time dimension; it flows over time.

We speak of utility per unit of time, productivity per unit of time, and so on. We do not speak of a commodity per unit of time or of a capital good per unit of time. A stock is condensed flow and being condensed it is packed, as it were, into a moment of time. It has thus no time dimension. Shall we, then, say that, in a sense, a stock corresponds to the static concept in economics?

CHAPTER 19

Closed and Open Systems

MEANINGS

A THING is said to be closed when nothing can enter it and nothing can get out of it. When exit and entry to a thing are closed it is said to be closed. When a thing is not closed in this sense it is said to be open. An economic system is said to be a closed one when no exogenous influences can affect its operations and likewise when it, in its turn, cannot influence the operation of other systems. A closed system is, therefore, a self-contained and self-sustained system. Robinson Crusoe's economy was a closed economic system.

A very pertinent question may be asked here. Is there any possibility, even theoretical, of a system being effectively closed to all external influences? The answer to this question depends on what precisely we understand by external influences. Every economy, for instance, is affected by such external forces as rainfall, floods, earthquakes, etc. Heavy rainfall is an external influence and so are floods and earthquakes. If we have to use the concept of a closed system we must therefore limit the connotation of the word *closed*. By a closed economic system let us understand a system, then, that is free from the influence of human factors belonging to other systems. This is theoretically possible though it may not be possible to find today a system that is closed even in this limited sense.

For purposes of analysis it matters little whether a system is free from external human influences or not. Whenever we want we can

just ignore such influences and study an economic system to find out how it would have behaved in case it was a closed system. This makes our study partake of the nature of investigations carried on in the laboratory of a physical scientist. He can in his laboratory insulate his system from all exogenous forces. We cannot experiment in that way but we can ignore such forces.

Closed and Open Systems in Economics

Economists have studied both closed and open systems. Their theories and propositions were sometimes based on the assumption that the economic system was a closed one and sometimes on the assumption that it was an open one. In a rough way we can say that the study of a firm is a study of an open system while that of an industry is a study of a (comparatively) closed one. If we pass on from an industry to the whole country the study becomes more truly one of a closed system. However, to make it a perfectly closed system we have to assume that there is no economic tie with other countries. For instance, we must take the case of a country that does not trade with foreign countries, does not borrow from or lend them money, and has no other economic links with them.

Macroeconomic studies approximate more closely to those of closed systems than microeconomic studies. Perfect or pure competition relates to systems that are open, monopolistic competition to those that are partially closed and partially open and monopoly to a closed system (if it is a monopoly in the correct sense of the word).

In the study of the problems of trade cycles, when we construct models, we generally begin with a closed system and gradually throw it open to exogenous influences. The foreign influence that we thus allow to modify our models is most often that of foreign trade. In the study of international trade we are of necessity concerned with open systems.

Fundamental Characteristics of an Open System

We have observed that an open system is one which is subject to external influences emanating from human agencies. But the essential characteristic of an open system is that it reacts to such external influences and, by so reacting, modifies them quantitatively and, at times, qualitatively as well. In the study of an open system full account has therefore to be taken of the chain of action and reaction as between a system and its environment. When this is not done the system, though

apparently open, becomes, in its essentials, a closed one.

In an open system, therefore, equilibrium is not reached all at once. The exogenous forces and the endogenous ones keep on reacting with each other; the battle goes on till, ultimately, a peaceful solution is reached. In the case of a closed system, taken as a whole, equilibrium is reached in a more direct fashion. Take for example the case of Crusoe. If there is any change in his wants or in any variable of his endogenous system the various forces readjust themselves to the new situation. There is a sort of willing co-operation among the various elements and a determination to reach a peaceful solution. The only opposition that Crusoe's economy might encounter would come from the side of natural forces. There are no forces other than the forces of nature, no human beings, to interfere with Crusoe's efforts to re-establish equilibrium. If Crusoe's want for fruits gets intensified he plucks more fruits. That will take more time and there will be a greater expenditure of energy. But there is willingness to make these needed sacrifices. Equilibrium is reached without much delay. If, however, Crusoe was to get more fruits by an act of bartering his goods with a fruit-seller, equilibrium would take a longer time to be reached. For, the fruit-seller would try to exploit the situation created by Crusoe's intensified want for fruits.

Crusoe's economy is a closed system in all its essentials. The macro-economy of a country is not so perfectly closed. One can imagine the entire population of a country to constitute an indivisible whole but in actual practice inter-individual struggle goes on in a country's economy, making the process of reaching equilibrium subject to the same delay as one witnesses in the case of an open system.

A Closed System in the Theory of Consumption

In the theory of consumption we study how an individual or a group of persons behave in order directly to satisfy wants. For the sake of simplicity, let us talk in terms of an individual. He has wants and he has means to satisfy these wants fully or partially. In order to satisfy his wants he has to depend on his environment. If he is living in a society, his environment consists of that society. He will then use his means to induce society to help him to get his wants satisfied. Suppose his want for a particular commodity, say, wheat increases. He will then need to buy more wheat in the market, offering a proportionately larger sum of money to induce the market to part with a larger quantity. It is assumed that his demand for wheat constitutes a small part of the total or market demand, and when this is so he can, without difficulty, buy more wheat by offering a proportionately larger amount of money. Such an individual or such a consumer, then, constitutes a closed system as he

hardly encounters resistance from his environment to his efforts to satisfy his want. His is a case similar to that of Crusoe who has to work harder and spend more time when his want for fruits increases.

In the theory of consumption, whether we use Marshallian demand-curve technique or the indifference-curve technique, we are mostly concerned with a closed system. But this is true only of those cases in which our consumer, by his increased demand, has an insignificant impact on the environment on which he depends for the satisfaction of his want. Since the impact is negligible the environment does not react to (oppose) the consumer's efforts to satisfy his intensified want. The only opposition that he encounters takes the shape of a demand for a proportionately larger sum of money from the consumer.

Where, however, a consumer's demand does not constitute an insignificant part of total demand the system tends to open itself out to exogenous influences, the system tends to become an open one.

Duopoly: An Open System

In duopoly there are two sellers with a common group of buyers to sell their products to. Buyers are common in the sense that they are not attached to either of the sellers. They buy from him who charges a lower price. We assume that the products of the duopolists are identical in every respect — the physical commodity and the manner of selling it and the circumstances in which it is sold make identical appeals to the buyers. Each duopolist, then, constitutes an open system. He is constantly subject to the opposition of his rival and, in his turn, offers a similar opposition to him. Equilibrium of price and, therefore, the equilibrium of the system would take a long time to reach. Theoretically, perhaps, the time required to reach equilibrium would be infinitely long if, and so long as, the duopolists continue to be duopolists in the right sense of the term.

The essential point of difference between open and closed systems is one that needs to be remembered. In the case of an open system, theoretically equilibrating adjustments take an inordinately long time. In some cases, where certain behaviouristic assumptions are made, equilibrium is reached only at infinity.

Augustin Cournot gave us a solution of the price problem in the case of duopoly where products are identical. In solving the problem it was assumed that each duopolist, while determining his own output, assumes that it will not affect the output of his rival. This assumption (or, as we shall see, some similar assumption) has to be made for a mathematical solution of the duopoly problem. Each duopolist sells that amount which maximises his profit (net income) under the assump-

tion that his action will not alter the output of the other seller which, therefore, is treated as a parameter. In this way the open system becomes virtually a closed system. But the system continues in reality to be open — the price continues to fluctuate and also the outputs — but at every step each duopolist thinks that the system is a closed one. This reveals stupidity on the part of the sellers. And since they cannot be assumed to be stupid we have to admit that the assumption made regarding the behaviour of sellers is itself stupid.

In any case, under such a stupid assumption a solution is reached. It gives outputs for the sellers which will stay and neither seller will find it to his advantage to change them. But this soution is yielded by a very unrealistic assumption. If and when a solution is reached and equilibrium is established (under assumptions which are so unrealistic), interaction between the duopolists comes to an end. The system of each seller is metamorphosed into a closed one. Equilibrium thus appears to be associated only with a closed system.

Edgeworth solves the problem of duopoly by assuming that each duopolist while fixing his price thinks it will not induce the rival to change his price. The price of the rival is taken to be a parameter. To that extent (in regard to price) the system is treated as a closed one. This enables Edgeworth to arrive at a solution of the price problem. But the price is found to be fluctuating between two limits, the upper one being close to monopoly price and the lower to competitive price. This is a solution, but it is a solution that yields no equilibrium price.

As in the case of Cournot's solution, here too the assumption reveals stupidity on the part of sellers. We are, however, concerned with another point, namely, that there is no equilibrium price. Both Cournot's and Edgeworth's solutions are artificial inasmuch as they rest on unrealistic assumptions regarding the behaviour of duopolists. As A.C. Pigou said, we cannot proceed without making some kind of assumption but there is no realistic assumption that would yield an equilibrium solution. *We repeat, a system has to be closed if there has to be equilibrium.*

Other economists have offered other solutions to the problem of duopoly. We are not concerned with these problems as such: our object is merely to use these solutions to show that equilibrium is not possible (at any rate not easily possible) in the case of an open system.

It has been shown that, in some cases, when the two duopolists are not of equal status, the one that has a superior status is able to dictate his price to the other seller. He becomes the leader and the other the follower. It is not always necessary for the two sellers to have different status. It is sufficient if they think that they can act as leaders. Leadership may be in regard to price or output. Heinrich von Stackelberg works out cases in which the leader sets the output and the follower adjusts his position, taking that output to be the leader's final decision.

Various varieties of such leadership are possible. An interesting position of disequilibrium results when each duopolist acts as a leader, thinking that the other would behave as a follower.

Here too one finds that for a solution to be arrived at, i.e., if equilibrium has to be achieved, the system must be a closed one. Where it is seemingly an open system it must virtually become a closed one if there has to be equilibrium. This is the burden of our song.

Monopoly—A Closed System

Monopoly is a closed system. A monopolist has not to face any opposition from his environment, if he is a monopolist in the right sense of the word. The opposition he meets with is, at the most, of the same nature as the opposition that Crusoe has to face when he uses his environment to satisfy his wants. A proper understanding of this point requires a correct knowledge of the meaning of monopoly.

The word monopoly has been variously defined; some of these definitions are theoretically unsound. But even a correct definition can be worded in more than one way. It is said that perfect competition (more correctly, pure competition) is characterised by the horizontality of the demand curve of a seller. Or, to make it more rigourous, one can say that, to the extent to which the demand curve for the product of a seller is horizontal, he can be regarded as belonging to a group of competing sellers. From this it should follow that a monopolist is one the demand curve for whose product is vertical. Again, to make the definition logically perfect, we can say that to the extent to which the demand curve is vertical the seller can be regarded as a monopolist.

Diagram 19.1 shows some mathematically possible demand curves. Figure *A* has a horizontal demand curve which depicts the picture of a producer when there is pure competition. Figure *B* has a vertical demand curve which depicts the picture of a (pure) monopolist. Figure *C* has a normal, negatively inclined, demand curve that depicts the picture of a market demand curve or, for the sake of comparison with the above cases, we might say that it depicts the case of a seller who is partly a monopolist and partly not. To make this point clear we have drawn in Figure *D* a curve which is made up of vertical and horizontal bits. If we were to examine the curve of Figure *C* through a powerful microscope it would look somewhat like the curve in Figure *D*. And this curve shows that the seller is a monopolist over vertical ranges of the demand curve but faces pure competition over horizontal ranges.

Our object here is not to write on the theory of monopoly and so we need not dilate on this point nor clarify some side issues that the concept of monopoly naturally raises, such as the meaning or definition

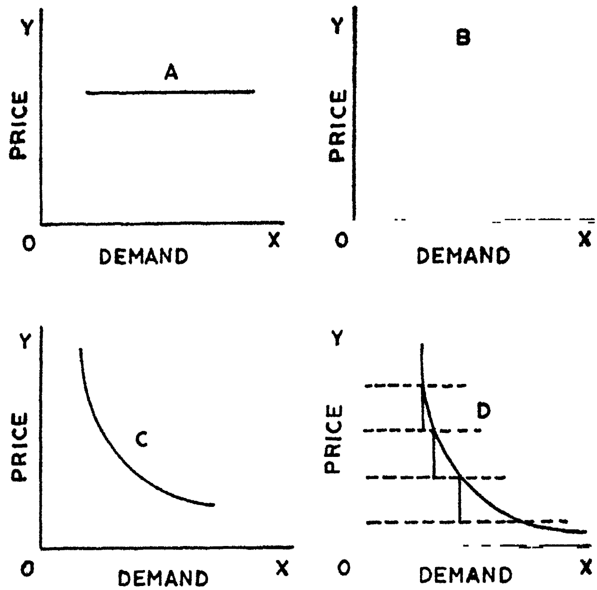


DIAGRAM 19.1

of a commodity. We simply concentrate our attention on the fact (which we accept) that a monopolist has a vertical demand curve for his product. Or, so long as the curve is vertical, the seller is a monopolist.

Where a seller is a monopolist, it follows from the nature of the demand curve, he can charge any price he likes (limited only by the range of the vertical demand curve) and yet sell the amount indicated by the demand curve. When a monopolist has such a hold on the buyers he has virtually no opposition to face from his environment. The only obstacle in his way is the sacrifice he has to make in producing or offering for sale the commodity he trades in. He is then in the same position in which Crusoe is, at any rate for all practical purposes. The monopolist's system is, therefore, a closed one. Equilibrium of the mind of the seller is readily attainable in such a case. He is at liberty to charge the price he likes to, at any moment of time. We may repeat here that, for equilibrium, a closed system is necessary. Where the actual situation is a mixture of open and closed systems, as it generally (perhaps always) is, equilibrium is only temporarily or partially attained.

MACROECONOMICS: CLOSED AND OPEN SYSTEMS

The economy of a country is a macroeconomy, if it is considered as in reality the *economy of a country* and not as an aggregate of the economies

of its various constituents. The trouble, as we have hinted elsewhere, is that no where do the various parts of a whole economy combine harmoniously to create a macroecomony in the true sense of the word. A macroeconomy must be a unitary concept: there should be just one unit, functioning as one unit. An individual, considered by himself and as isolated from other individuals, makes a true picture of macroeconomy. A country, on the other hand, has a large number of individuals and a large number of production and consumption units. There is conflict and struggle among them so that, when viewed as a whole, they merely constitute an aggregate of so many units. They do not merge themselves into one unit as they should if a true macroeconomy has to emerge out of them.

One can, by shutting one's eyes to facts, take a country's economy to be a macroeconomy. In building models of a macroeconomy, growth models and the like, one just concentrates on certain important and essential aggregates of variables and proceeds on the supposition that a true macroeconomy exists. If an economy is a macroeconomy in the true sense it functions as a unit; but our economies cannot so function. But if and when a macroeconomy exists it is a closed system inasmuch as it is free from the influence of all forces that are in any sense external to it. Crusoe economy is the most appropriate example. Otherwise, a country's economy is not a macroeconomy and is, therefore, an open system. And as we have observed above such an economy cannot attain equilibrium position. No wonder then that no country ever finds its economy in the position of equilibrium.

A country's economy that is already an open one becomes more fully open when it begins to trade with other countries, buying and selling goods and borrowing or lending money. Interaction between countries then starts and each party finds itself facing opposition from the other. Equilibrium becomes as difficult or impossible as in the case of duopoly or oligopoly. What we are witnessing today bears testimony to our thesis. There will never be any equilibrium in any country of the world so long as the economic system remains open. There is no world organisation today, nor will there ever be one, that combines the various countries into a real macroeconomy.

In the endogenous models of growth that exhibit fluctuations round a rising trend we introduce an exogenous factor sometimes by including the foreign trade variable into it. We make exports, at times, an autonomous investment-factor and imports an induced savings-factor. In this way we expose our system to shocks from the exogenous, external world. This is how we open our system and invite foreign influences to disturb it in its abortive attempt to reach the position of equilibrium.

CHAPTER 20

The Multiplier

Introductory

In mathematics the word 'multiplier' refers to the quantity which multiplies another quantity to give a result. Thus we get 12 (result) when we multiply 3 by 4: here the number 4 is a *multiplier*. Similarly, if 3 stands for a cause and 12 for the result, we may say that the cause leads to an effect which is 4 times the cause, or, that 4 indicates the multiplier action of cause on effect. Or, again, if in an economy we find that an initial employment of one more labourer results ultimately in an employment of 4 labourers, we may say that in that economy the multiplier action is 4. If we spend or invest one more rupee and find that it leads ultimately to transactions worth 4 rupees, we may say that the multiplier action of investment on transactions (during a given period of time) is indicated by 4.

In economics, the ratio of an initial increase in a factor to the resultant increase in income is called the *specific factor multiplier* (with reference to income). Thus we have

(1) Specific factor multiplier:
- (*i*) investment multiplier
- (*ii*) consumption multiplier
- (*iii*) export multiplier
- (*iv*) foreign investment multiplier
- (*v*) tax multiplier
- (*vi*) government expenditure multiplier

(*vii*) transfer payments multiplier
(*viii*) balanced budget multiplier
(2) simple and compound multiplier
(3) curvilinear multiplier
(4) multi-effect multiplier
(5) multi-sector multiplier
(6) static and dynamic multiplier:
(a) multiplier with Robertsonian lag
(b) multiplier with Lundbergian lag
(c) multiplier with consumption and production lags

Now there are two ways of studying and analysing these multipliers. One, we may collect observations about income and the factor concerned and fit a relationship between them statistically by (say) the method of least squares

$$\begin{aligned}
\text{Income} &= a_1 + k_1 \text{ Investment}\\
\text{Income} &= a_2 + k_2 \text{ Tax}\\
\text{Income} &= a_3 + k_3 \text{ Government expenditure}\\
\text{Income} &= a_4 + k_4 \text{ Export}
\end{aligned}$$

and calculate the values of a's and k's.
Symbolically

$$\begin{aligned}
Y &= a_1 + k_1\ I\\
Y &= a_2 + k_2\ T\\
Y &= a_3 + k_3\ G\\
Y &= a_4 + k_4\ E
\end{aligned}$$

Then k_1, k_2, k_3 and k_4 will be the respective multipliers.

We may have more complex relationships between Y and the factor concerned. Thus we may write

$$Y = a_1 + k_{11}\ I + k_{12}\ I^2$$

Or, we may have more than one factor involved in the same equation:

$$Y = a_1 + k_1\ I + k_2\ T + k_3 G$$

Statistical methods would then enable us to find the value of k_1, k_2 and k_3 which will then be the values of the multipliers.

This would be an inductive way but we have to decide what should be the form of relationship to be fitted. Hence we may go on to the deductive side and examine how income is related to these factors or variables. We may then theoretically analyse the expressions for multipliers through the use of differential calculus or the calculus of finite variation.

We know that income or output(Y) is divided into two major parts — consumption (C) and investment(I), so that

$$Y = C + I$$

If a part of the income is used up by the Government also, we write

$$Y = C + I + G$$

If a part is exported — export being taken in the sense of net export (E) — we write:

$$Y = C + I + G + E$$

or

$$Y = C + I + G + (X - M)$$

where X indicates exports and M imports.

Now consumption can be looked upon as *dependent* on income (Y) and tax (T) paid

$$C = \alpha + c\ (Y - T)$$

i.e., consumption is dependent on income left after paying taxes. We may instead say "after paying taxes (T) and receiving transfer payments (T_r) from government": in which case

$$C = \alpha + c\ (Y - T + T_r)$$

and if we put this value of C we can rewrite

$$Y = C + I + G + (X - M)$$

as

$$Y = \alpha + c(Y - T + T_r) + I + G + (X - M)$$

or

$$Y(1 - c) = \alpha - cT + cT_r + I + G + (X - M)$$

To understand the multiplier due to (say) I, we imagine a small increase in I and a consequent small increase in Y. We indicate these small increases by ΔI and ΔY respectively. Then we can write

$$(Y + \Delta Y)\ (1 - c) = \alpha - cT + cT_r + (I + \Delta I) + G + (X - M)$$

Hence the difference between the two equations can be equated, giving us

$$(\Delta Y)\ (1 - c) = \Delta I$$

$$\therefore\ \frac{\Delta Y}{\Delta I} = \frac{1}{1 - c}$$

So $\frac{1}{1-c}$ indicates the ratio of the increase in Y per unit increase in I and is the multiplier-effect of I on Y. It is called the investment multiplier, k_i:

$$k_i = \frac{1}{1 - c}$$

Similarly, if we denote the tax multiplier by k_T, we can get

$$k_T = \frac{\Delta Y}{\Delta T} = -\frac{c}{1 - c}$$

For others, we have

$$k_G = \frac{\Delta Y}{\Delta G} = \frac{1}{1 - c}\ \text{(Government expenditure multiplier)}$$

$$k_{Tr} = \frac{\Delta Y}{\Delta T_r} = \frac{c}{1-c} \text{ (Transfer payment multiplier)}$$

$$k_X = \frac{\Delta Y}{\Delta X} = \frac{1}{1-c} \text{ (Export multiplier)}$$

$$k_M = \frac{\Delta Y}{\Delta M} = -\frac{1}{1-c} \text{ (Import multiplier)}$$

If we had used E for net export, we could get

$$k_E = \frac{\Delta Y}{\Delta E} = \frac{1}{1-c}$$

If we had assumed that government expenditure is equal to taxes levied (net of transfer payments) then, instead of $(T - T_r)$ we would write G:

$$Y = \alpha + c\,(Y - G) + I + G + (X - M)$$

and get $\frac{\Delta Y}{\Delta G} = \frac{1-c}{1-c} = 1 =$ (say) k_{BB}, where k_{BB} is termed the *balanced budget multiplier*: it is always unity.

When we write

$$Y = \alpha + cY + I + G + E$$

we seem to assume that E, G and I are determined independently. However, these may be assumed to be interconnected or dependent on income also. Thus *if* I includes government investment determined by (say) $\frac{\alpha_1}{Y}$, we may explicitly write

$$Y = \alpha + cY + \left(I + \frac{\alpha_1}{Y}\right) + G + E$$

in which case we will find that the investment multiplier is a complicated expression. Now

$$Y + \Delta Y = \alpha + c\,(Y + \Delta Y) + \left[(I + \Delta I) + \frac{\alpha_1}{Y + \Delta Y}\right] + G + E$$

and therefore, subtracting from this the original Y-equation given above, we get

$$\Delta Y = c(\Delta Y) + \Delta I + \alpha_1\left[\frac{1}{Y + \Delta Y} - \frac{1}{Y}\right]$$

$$= c(\Delta Y) + \Delta I + \alpha_1\left[-\frac{\Delta Y}{Y\,(Y + \Delta Y)}\right]$$

or,

$$= c(\Delta Y) + \Delta I - \frac{\alpha_1 \Delta Y}{Y^2}\left[\frac{1}{1 + \frac{\Delta Y}{Y}}\right]$$

$$= c(\Delta Y) + \Delta I - \frac{\alpha_1 \Delta Y}{Y^2}\left[1 - \frac{\Delta Y}{Y}\right] \text{ approximately.}$$

If we ignore $\frac{\Delta Y}{Y}$ of $1 - \frac{\Delta Y}{Y}$, we get

$$\Delta Y = c(\Delta Y) + \Delta I - \frac{\alpha_1 \Delta Y}{Y^2}$$

$$\therefore \qquad \left(1 - c + \frac{\alpha_1}{Y^2}\right) \Delta Y = \Delta I$$

$$\therefore \qquad \frac{\Delta Y}{\Delta I} = \frac{1}{1 - c + \frac{\alpha_1}{Y^2}}$$

which will then be the new investment-multiplier.

Similarly, consumption may not be expressed as

$$C = \alpha + cY$$

as it means that, when income increases infinitely, consumption will also increase infinitely. We may imagine that (*i*) the coefficient c of Y will itself decrease as Y increases, or (*ii*) that there is a ceiling on C itself by imagining it as given by

$$C = \frac{\alpha}{1 + \beta c^{-Y}}$$

So that if Y becomes infinitely large, c^{-Y} becomes zero and we get a maximum consumption

$$C = \alpha$$

The use of this expression for C in the identity

$$Y = C + I + G + E$$

will no doubt complicate the analysis and the determination of the 'multiplier': but it gives an indication of new directions which may be pursued in modern economics.

Similarly, the identity relationship being

$$\begin{aligned} Y &= C + I + G + X - M \\ &= \{\alpha + c(Y - T + T_r)\} + I + G + X - M \end{aligned}$$

We may write

$$\Delta Y = [\{c(\Delta Y - \Delta T + \Delta T_r)\} + \Delta I + \Delta G + \Delta X - \Delta M] + 1$$

where '1' stands for an autonomous increase in any of the right-hand components, and if *then* we express

$$\begin{aligned} \Delta T &= z.\Delta Y \\ \Delta T_r &= z_r.\Delta Y \\ \Delta I &= \beta.\ \Delta Y \\ \Delta G &= g.\ \Delta Y \\ \Delta X &= e.\ \Delta Y \\ \Delta M &= m.\Delta Y \end{aligned}$$

where z = marginal propensity to tax
z_r = marginal propensity to transfer
β = marginal propensity to invest
g = marginal propensity of government expenditure
e = marginal propensity to export
m = marginal propensity to import

We will find that

$$\Delta Y = \frac{1}{1 - c + cz - cz_r - \beta - g - e + m}$$

If, instead, we want to find the multiplier for investment, we may not write $\beta . \Delta Y$ but retain ΔI on the right-hand side as an autonomous increase instead of '1'. We will then get

$$\Delta Y = \frac{\Delta I}{1 - c + cz - cz_r - g - e + m}$$

i.e., the corresponding marginal propensity term vanishes from the denominator to give us the relevant multiplier. Thus seen,

$$\frac{\Delta Y}{\Delta X} = \frac{1}{1 - c + cz - cz_r - \beta - g + m} = \text{Export multiplier}$$

$$\frac{\Delta Y}{\Delta M} = \frac{1}{1 - c + cz - cz_r - \beta - g - e} = \text{Import multiplier}$$

If $E = X - M$ so that $\Delta E = \Delta X - \Delta M$, we can write

$$\frac{\Delta Y}{\Delta E} = \frac{1}{1 - c + cz - cz_r - \beta - g} = \text{Net export multiplier}$$

$$\frac{\Delta Y}{\Delta G} = \frac{1}{1 - c + cz - cz_r - \beta - e + m} = \text{Government expenditure multiplier}$$

If, in connection with the government expenditure multiplier, we were to impose the condition $T = G$ and $\Delta T = \Delta G$, we will not write $\Delta T = z . \Delta Y$ but write $\Delta T = \Delta G$ and get

$$\Delta Y = c(\Delta Y - \Delta G) + \Delta I + \Delta G + \Delta X - \Delta M$$

$$\frac{\Delta Y}{\Delta G} = \frac{1 - c}{1 - c - \beta - e + m}$$

One can think of many other improvements, modifications and changes before deriving an expression for a certain multiplier.

Simple and Compound Multiplier

We know that, as a simple exposition,

$$Y = C + I$$

If $C = \alpha + \beta Y$, and I changes autonomously from I_1 to I_2 we can write

$$Y_1 = C_1 + I_1 = \alpha + cY_1 + I_1$$

$$Y_2 = C_2 + I_2 = \alpha + cY_2 + I_2$$

$$\therefore \quad Y_2 - Y_1 = c(Y_2 - Y_1) + I_2 - I_1$$

or

$$\frac{Y_2 - Y_1}{I_2 - I_1} = \frac{1}{1 - c}$$

So the multiplier is $\frac{1}{1-c}$ and it is called a simple multiplier.

In $C = \alpha + cY$, cY indicates consumption as dependent on income and α indicates autonomous consumption.

If investment (I) had not changed but α had changed from α_1 to α_2, then we would have

$$Y_1 = C_1 + I = \alpha_1 + cY_1 + I$$
$$Y_2 = C_2 + I = \alpha_2 + cY_2 + I$$
$$\therefore \quad Y_2 - Y_1 = (\alpha_2 - \alpha_1) + c(Y_2 - Y_1)$$
$$\text{or} \quad \frac{Y_2 - Y_1}{\alpha_2 - \alpha_1} = \frac{1}{1-c}$$

Now $\frac{1}{1-c}$ will be the multiplier obtained as a result of an autonomous change in consumption.

Similarly, if we wrote

$$Y = C + I + G + E = \alpha + cY + I + G + E$$

and government expenditure, G, or export, E, were to change autonomously we would get

$$\frac{Y_2 - Y_1}{G_2 - G_1} = \frac{1}{1-c}$$
$$\frac{Y_2 - Y_1}{E_2 - E_1} = \frac{1}{1-c}$$

The multiplier $\frac{1}{1-c}$ can therefore be called

(*i*) *investment multiplier*, due to an autonomous change in investment;

(*ii*) *consumption multiplier*, due to an autonomous change in consumption;

(*iii*) *government expenditure multiplier*, due to an autonomous change in government expenditure; or

(*iv*) *export multiplier*, due to an autonomous change in net export.

These may be said to be cases of *simple multiplier* because nothing but consumption alone is taken to depend on income also.

If we take investment also to depend on income according to the formula

$$I = \gamma + \beta Y$$
$$Y = C + I = \alpha + cY + \gamma + \beta Y$$
$$\therefore \quad Y_1 = \alpha + cY_1 + \gamma_1 + \beta Y_1$$
$$Y_2 = \alpha + cY_2 + \gamma_2 + \beta Y_2$$
$$\therefore \quad Y_2 - Y_1 = c(Y_2 - Y_1) + (\gamma_2 - \gamma_1) + \beta(Y_2 - Y_1)$$
$$\text{or} \quad \frac{Y_2 - Y_1}{\gamma_2 - \gamma_1} = \frac{1}{1 - c - \beta}$$

This multiplier is called a *compound multiplier* because in deriving its value we assume that more than one component depends on income.

Curvilinear Multiplier

We have so far assumed a linear function for consumption: it may instead be non-linear, such as,

$$C = \alpha + cY - c'Y^2$$

in which case, we would get as a result of an autonomous increase in I

$$Y_2 - Y_1 = c(Y_2 - Y_1) - c'(Y_2^2 - Y_1^2) + (I_2 - I_1)$$

$$\therefore \quad \frac{Y_2 - Y_1}{I_2 - I_1} = \frac{1}{1 - c + c'(Y_2 + Y_1)}$$

We may call such a multiplier a *curvilinear multiplier.*

If, similarly, the investment function were also taken to be

$$I = \gamma + \beta Y + \beta' Y^2$$

we would have

$$\begin{aligned} Y_2 - Y_1 &= (C_2 - C_1) + (I_2 - I_1) \\ &= c(Y_2 - Y_1) - c'(Y_2^2 - Y_1^2) + (\gamma_2 - \gamma_1) + \beta(Y_2 - Y_1) + \beta'(Y_2^2 - Y_1^2) \end{aligned}$$

$$\therefore \quad \frac{Y_2 - Y_1}{\gamma_2 - \gamma_1} = \frac{1}{1 - c - \beta + (c' - \beta')(Y_2 + Y_1)}$$

Multi-effect Multiplier

We may well imagine a simultaneous autonomous change in both consumption and investment and study their combined effect on income increase. Thus, for a simple case of the type

$$C_1 = \alpha_1 + cY_1$$

$$C_2 = \alpha_2 + cY_2$$

we will have

$$Y_1 = C_1 + I_1 = \alpha_1 + cY_1 + I_1$$

$$Y_2 = C_2 + I_2 = \alpha_2 + cY_2 + I_2$$

$$\therefore \quad Y_2 - Y_1 = (\alpha_2 - \alpha_1) + c(Y_2 - Y_1) + I_2 - I_1$$

or, $$(1 - c)(Y_2 - Y_1) = (\alpha_2 - \alpha_1) + (I_2 - I_1)$$

or, $$\frac{Y_2 - Y_1}{(\alpha_2 - \alpha_1) + (I_2 - I_1)} = \frac{1}{1 - c}$$

One can obtain similar results with more complex functions for C and I and for more detailed break-up of Y than into C and I only.

Multi-sector Multiplier

There may even be interdependence between C and I and the economic activities may be conceived as represented in a social accounting table or input-output table. In this case an autonomous change (*i*) in investment in any industry (or sector) or (*ii*) in expenditure will show up as a change in many sectors; and in the equilibrium position the entries in the table may be totally different. The ratio of income-increase to the autonomous change may be called a *multi-sector multiplier.* An alternative name may be *matrix-multiplier.* Experts like Goodwin (R.M.) and Turvey (R.) have presented studies in these directions in the pages of the *Economic Journal* (1949) and the *American Economic Review* (1953).

Static and Dynamic Multiplier

The various types of multiplier mentioned above have one basic feature. We have jumped from one equilibrium position to another and studied the ratio of the rise in income to autonomous rise in a component. We have not studied how the new equilibrium position is reached. Therefore, the exposition has at times been termed a study in a *static multiplier,* though it would perhaps be better to call it a static study of the multiplier-effect.

On the other hand, we might study how the new equilibrium position is reached, i.e., how the multiplier-effect materialises in a simplified version of actual life. When this is done, it exposes the dynamic multiplier aspect and the study may be called a *dynamic study of the multiplier-effect*: it has sometimes been termed the *dynamic multiplier.*

Multiplier with Robertsonian Lag

We may, for the purpose of getting a slightly more detailed idea of the so-called dynamic multiplier, assume that

$$C_t = \alpha + cY_{t-1}$$

i.e., there is a lag—called the *Robertsonian Lag* after Robertson (Sir D.) who mentioned it in 1926 in his study *Banking Policy and the Price Level*—of one period in consumption. Then

$$Y_t = C_t + I_t = \alpha + cY_{t-1} + I_t$$

$$Y_{t+1} = C_{t+1} + I_{t+1} = \alpha + cY_t + I_{t+1}$$

$$\therefore \quad Y_{t+1} - Y_t = c(Y_t - Y_{t-1}) + (I_{t+1} - I_t)$$

or $$\Delta Y_t = c\Delta Y_{t-1} + \Delta I_t$$

where $Y_{t+1} - Y_t = \Delta Y_t$; hence $Y_t - Y_{t-1} = \Delta Y_{t-1}$ and $I_{t+1} - I_t = \Delta I_t$

$$\therefore \quad \frac{\Delta Y_t}{\Delta I_t} = c.\frac{\Delta Y_{t-1}}{\Delta I_t} + 1 = c.\frac{\Delta Y_{t-1}}{\Delta I_{t-1}}.\frac{\Delta I_{t-1}}{\Delta I_t} + 1$$

If we use k_t to indicate the ratio or multiplier $\frac{\Delta Y_t}{\Delta I_t}$ we can write

$$\begin{aligned} k_t &= c \cdot \frac{\Delta I_{t-1}}{\Delta I_t} \cdot k_{t-1} + 1 \\ &= c \cdot \frac{\Delta I_{t-1}}{\Delta I_t} \cdot k_{t-1} \text{ approximately (on ignoring 1)} \\ &= \left(c.\frac{\Delta I_{t-1}}{\Delta I_t}\right)\left(c \cdot \frac{\Delta I_{t-2}}{\Delta I_{t-1}}\right) \cdot k_{t-2} \end{aligned}$$

and ultimately $$= c^{t-1}.\frac{\Delta I_{t-1}}{\Delta I_t}.\frac{\Delta I_{t-2}}{\Delta I_{t-1}} \ldots\ldots \frac{\Delta I_2}{\Delta I_3}.\frac{\Delta I_1}{\Delta I_2}.k_1$$

$$= c^{t-1}.\frac{\Delta I_1}{\Delta I_t}.k_1$$

This may be said to be an exposition of the dynamic multiplier.

If we had conceived of an autonomous increase in consumption, and not in investment, we would get

$$k_t = c^{t-1} \cdot \frac{\Delta \alpha_1}{\Delta \alpha_t} \cdot k_1$$

Multiplier with Lundbergian Lag

Similarly, one may introduce on the production side a production time-lag—called the *Lundbergian Lag* after Lundberg (E.) for its mention in his *Studies in the Theory of Economic Expansion.* If we assume (*i*) that excess demand is met out of stock, and (*ii*) that investment is equal to an autonomous part plus another part to make up the depletion in stock experienced during the last period, and (*iii*) that it takes one year for output (Y_t) to catch up with total demand (D_{t-1}) we may write—

$$C_t = cY_t = cD_{t-1}$$
$$I_t = \gamma_t + \beta\,(D_{t-1} - D_{t-2})$$

because we invest to make up the fall in stock for meeting the excess demand in the last period. Replacing all D's by Y's we have

$$I_t = \gamma_t + \beta\,(Y_t - Y_{t-1})$$
$$\therefore \quad Y_t = C_t + I_t = cY_t + \gamma_t + \beta\,(Y_t - Y_{t-1})$$
$$\therefore \quad (1 - c - \beta)\,Y_t = \gamma_t - \beta Y_{t-1}$$

and

$$\therefore \quad (1-c-\beta)\, Y_{t+1} = \gamma_{t+1} - \beta Y_t$$

$$\therefore \quad (1-c-\beta)\, \Delta Y_t = \Delta\gamma_t - \beta \Delta Y_{t-1}$$

$$\therefore \quad (1-c-\beta) \cdot \frac{\Delta Y_t}{\Delta\gamma_t} = 1 - \beta \cdot \frac{\Delta Y_{t-1}}{\Delta\gamma_t}$$

$$= 1 - \beta \cdot \frac{\Delta Y_{t-1}}{\Delta\gamma_{t-1}} . \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t}$$

or, $$(1-c-\beta)\, k_t = 1 - \beta . k_{t-1} \cdot \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t}$$

$$\therefore \quad k_t = \frac{\beta}{\beta + c - 1} \cdot k_{t-1} \cdot \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t} + \frac{1}{1-c-\beta}$$

If we ignored $\dfrac{1}{1-c-\beta}$, we would write

$$k_t = \left(\frac{\beta}{\beta+c-1} \cdot \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t}\right) k_{t-1}$$

$$= \left(\frac{\beta}{\beta+c-1} \cdot \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t}\right)\left(\frac{\beta}{\beta+c-1} \cdot \frac{\Delta\gamma_{t-2}}{\Delta\gamma_{t-1}}\right) k_{t-2}$$

$$\text{and ultimately} = \left(\frac{\beta}{\beta+c-1}\right)^{t-1} \frac{\Delta\gamma_1}{\Delta\gamma_t} \cdot k_1$$

Multiplier with Consumption and Production Lags

We may even take account of both lags. Writing

$$C_t = \alpha + cY_{t-1}$$

$$I_t = \gamma_t + \beta(Y_t - Y_{t-1})$$

$$Y_t = \alpha + cY_{t-1} + \gamma_t + \beta(Y_t - Y_{t-1})$$

or, $$(1-\beta)\, Y_t = \alpha + (c-\beta)\, Y_{t-1} + \gamma_t$$

$$\therefore \quad (1-\beta)\, Y_{t+1} = \alpha + (c-\beta)\, Y_t + \gamma_{t+1}$$

$$\therefore \quad (1-\beta)\,(Y_{t+1} - Y_t) = (c-\beta)\,(Y_t - Y_{t-1}) + (\gamma_{t+1} - \gamma_t)$$

or, $$(1-\beta)\, \Delta Y_t = (c-\beta)\, \Delta Y_{t-1} + \Delta\gamma_t$$

$$\therefore \quad \frac{\Delta Y_t}{\Delta\gamma_t} = \frac{c-\beta}{1-\beta} \cdot \frac{\Delta Y_{t-1}}{\Delta\gamma_{t-1}} \cdot \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t} + 1$$

Ignoring 1, as before,

$$k_t = \frac{c-\beta}{1-\beta} \cdot \frac{\Delta\gamma_{t-1}}{\Delta\gamma_t} \cdot k_{t-1}$$

$$= \left(\frac{c-\beta}{1-\beta}\right)^{t-1} \frac{\Delta\gamma_1}{\Delta\gamma_t} \cdot k_1$$

These are mathematical expositions and need economic interpretations. The application of mathematics to economics tends to proceed not infrequently (as here) without such interpretation.

The 'multiplier' concept is useful only in so far as it helps us get an idea of how *ultimately* a change in a factor affects the increase in income or output of a society, provided the determination of the factor-value is independently (i.e., autonomously) done. Thus, where we have to choose between a number of factors (T, T_r, G, E) for action so as to affect income (Y), the different multipliers indicate their comparative values: between two factors, we would choose a factor having a greater multiplier-effect.

The multiplier may also be used to fix a maximum expansion ratio so as not to engender a balance of payments problem. Thus, if the marginal propensity to save is $\frac{1}{4}$, that of imports $\frac{1}{4}$, the total propensity to invest is $\frac{1}{2}$ and the multiplier is 2, given a certain rate of capital-inflow, there will be a certain maximum income-growth-ratio (or, a certain expansion-ratio) which designates the limit to which domestic investment can be increased. This is particularly important for developing poor countries when capital-inflow may lead to demand-creation in other industries leading to more domestic investment in such a way (say, through created money) as to add to the demand for imports to such an extent that the increased foreign exchange requirement exceeds the supply of foreign exchange.

Limitations of Multiplier Concept

The above presentation of multiplier-analysis does not make it clear whether the various values (Y, C, I and G) are expressed in real terms or in money terms. We *usually* take them as expressed in money terms and yet ignore the total quantity of money available. But implicitly we assume that (say) the extra investment (ΔI) is in the form of creation of extra money: otherwise, extra expenditure under (say) I or E or T will be accompanied by a decreased expenditure under some other item. If we think in terms of real resources, extra expenditure will be possible only if there are unused physical resources and manpower. Besides, in actual practice, why should extra investment be made? It will be made only (*i*) if investors (or entrepreneurs) expect extra demand or profit and (*ii*) if those who make extra money available believe in these expectations (otherwise, why should they risk their money?).

In actual life aggregate consumption depends not merely on income (whether preceding or current or both) but on many other factors such as distribution of income, population, prices, tastes, preferences and demonstration-effects. Similarly, investment depends on consi-

derations of income, rate of interest and expectations.

Besides, it is a weakness that we have to assume that when one determining factor changes (i.e., when graphically speaking, its behaviour equation shifts vertically) other factors do not change. For this assumption is questionable: thus, in regard to exports this assumption ignores the conditions obtaining in other countries (regions or sectors).

It may be added that there exist such concepts also as

(*i*) money-creation multiplier
(*ii*) deposit multiplier
(*iii*) employment multiplier

Money-creation Multiplier

If the banking system has the policy (or responsibility) to maintain a minimum reserve-ratio between its reserves and current liabilities, then, with an excess reserve of ΔZ, it can create an additional credit money (ΔK_r) given by

$$\Delta K_r = \left(\frac{1}{r}\right) \Delta Z$$

and $\frac{1}{r}$ is called the *money-creation multiplier.*

Deposit Multiplier

Let the people have a habit of drawing a proportion γ of the created money as cash and of depositing the rest. Let us indicate by ΔN the amount drawn and by ΔD the amount deposited out of the created money. Then we can write:

$$\Delta K_r = \Delta N + \Delta D \quad \text{(Identity equation)}$$
$$\Delta N = \gamma . \Delta K_r \quad \text{(Behaviour equation)}$$
$$(\Delta Z - \Delta N) = r . \Delta D \quad \text{(Equilibrium equation)}$$

It may be mentioned that ΔZ is excess reserve ex-ante: ex-post, the additional reserve in hand will be $\Delta Z - \Delta N$, i.e., ΔZ will be reduced by the additional cash (ΔN) drawn out by the public.

$$\begin{aligned} \therefore \quad \Delta K_r &= \Delta N + \Delta D \\ &= \Delta N + \frac{1}{r}(\Delta Z - \Delta N) \\ &= -\left(\frac{1}{r} - 1\right) \Delta N + \frac{1}{r} \cdot \Delta Z \\ &= -\left(\frac{1}{r} - 1\right) \cdot \gamma . \Delta K_r + \frac{1}{r} \cdot \Delta Z \end{aligned}$$

$$\therefore \quad \left[1 + \gamma\left(\frac{1}{r} - 1\right)\right] \Delta K_r = \frac{1}{r} \cdot \Delta Z$$

$$\text{or,} \quad [r + \gamma(1 - r)]\, \Delta K_r = \Delta Z$$

$$\therefore \quad \Delta K_r = \frac{1}{r + \gamma(1 - r)} \cdot \Delta Z$$

$$\begin{aligned} \therefore \quad \Delta D &= \Delta K_r - \Delta N \\ &= \Delta K_r - \gamma \cdot \Delta K_r \\ &= (1 - \gamma)\,.\, \Delta K_r \\ &= \frac{1 - \gamma}{r + \gamma(1 - r)} \cdot \Delta Z \end{aligned}$$

Hence $\dfrac{1 - \gamma}{r + \gamma(1 - r)}$ is called the *deposit multiplier.*

Similarly, if the banking system in equilibrium is able to get additional cash $\Delta Z'$ from the non-banking sector as deposit, it will have to retain $r.\Delta Z'$ as reserve and hence its excess reserve will be $\Delta Z' - r.\Delta Z'$. Hence the additional credit it can create will be given by

$$\Delta K_r = \frac{\Delta Z' - r.\Delta Z'}{r + \gamma(1 - r)} = \frac{1 - r}{r + \gamma(1 - r)} \cdot \Delta Z'$$

However, we can easily see that

(*i*) given γ, ΔK_r declines with increasing r, so much so that if $r = 1$, the multiplier becomes 1 whatever γ may be, and

(*ii*) given r, ΔK_r declines with increasing γ so much so that if $\gamma = 1$, the multiplier becomes 1 whatever r may be: this corresponds to the situation where the debtor withdraws his loan in cash. In this case the deposit-multiplier will be zero.

An excess reserve (or cash) can be created in the banking system in one of the following ways:

(*a*) The central bank of the country may acquire assets and thus make cash available to banks.

(*b*) The central bank may reduce the minimum reserve ratio, r.

(*c*) The banking system may, by granting a loan, acquire, in the shape of security, assets eligible for rediscount with the central bank.

(*d*) The non-banking sector may acquire a greater habit of making payments by cheques.

(*e*) The non-banking sector may draw (or keep) a smaller proportion of loans as cash.

(*f*) The non-banking sector may convert a part of current-deposits into time-deposits, thereby reducing the reserve-requirements of the banking system, as the minimum reserve ratio for time-deposits is lower than that for demand-deposits.

If we may determine γ for a country, we can study the credit-creation

capacity of the banking system there. It may be mentioned that, if not Υ, we may first determine for a country

$$\frac{\Delta N}{\Delta D}$$

i.e., the ratio (say, k) of the cash retained by the non-banking sector to its deposits and then calculate Υ for if

$$k = \frac{\Delta N}{\Delta D}$$

$$1 + k = 1 + \frac{\Delta N}{\Delta D} = \frac{\Delta N + \Delta D}{\Delta D}$$

and
$$\frac{k}{1+k} = \frac{\Delta N}{\Delta N + \Delta D} = \frac{\Delta N}{\Delta K_r} = \Upsilon$$

Employment Multiplier

Just as $\frac{Y_2 - Y_1}{I_2 - I_1}$, $\frac{\Delta Y}{\Delta I}$ or $\frac{dY}{dI}$ is taken to be a measure of investment multiplier, one *could* call $\frac{Y_2 - Y_1}{l_2 - l_1}$ as the employment multiplier where $l_2 - l_1$ indicates a marginal increase in employment. But conventionally, this is not the concept attached to the employment multiplier. In practice, the employment multiplier stands for the ratio of

(*i*) ultimate increase in employment, to
(*ii*) initial increase in employment.

If additional resources are being invested, the initial increase in employment will be in some capital goods industry; this will lead to further employment of labour in other capital goods and consumer goods industries. Hence the total increase in employment will be more than the initial increase in employment. We may say that

$$k_e = \frac{l_f - l_1}{l_2 - l_1}$$

where l_1 is the initial level of employment; $l_2 - l_1$, the initial increase in employment, and l_f, the final level of employment.

Assuming that labour-productivity (M) does not change, i.e., assuming that the output-labour ratio $\left(\frac{Y}{l}\right)$ is constant (say $= M$) we can say that

$$\frac{Y_1}{l_1} = \frac{Y_2}{l_2} = \frac{Y_f}{l_f} = M$$

or
$$l_1 = \frac{Y_1}{M};\ l_2 = \frac{Y_2}{M};\ l_f = \frac{Y_f}{M}$$

$$\therefore \qquad k_e = \frac{l_f - l_1}{l_2 - l_1} = \frac{Y_f - Y_1}{Y_2 - Y_1} = \frac{\Delta Y}{\Delta I} = \text{Investment multiplier, } k_I$$

if $Y_2 - Y_1$ will mean initial increase in investment and $Y_f - Y_1$ will mean the ultimate increase in income. In this way, the employment multiplier will be the same as the investment multiplier.

It is, however, not necessary that labour-productivity should remain constant. If labour-productivity increases, then

$$\frac{Y_f}{l_f}$$

will be more than M: say, it will be $M + m$. Then

$$K_e = \frac{\dfrac{Y_f}{M+m} - \dfrac{Y_1}{M}}{\dfrac{Y_2}{M} - \dfrac{Y_1}{M}}$$

$$= \frac{\dfrac{Y_f}{M}\left(1 - \dfrac{m}{M}\right) - \dfrac{Y_1}{M}}{\dfrac{Y_2}{M} - \dfrac{Y_1}{M}} \text{ approximately}$$

as, ignoring higher powers of $\frac{m}{M}$,

$$\frac{1}{M+m} = \frac{1}{M\left(1 + \dfrac{m}{M}\right)} = \frac{1 - \dfrac{m}{M}}{M} \text{ approximately}$$

$$\therefore \quad K_e = \frac{Y_f\left(1 - \dfrac{m}{M}\right) - Y_1}{Y_2 - Y_1}$$

$$= \frac{Y_f - Y_1}{Y_2 - Y_1} - \frac{m}{M}\,\frac{Y_f}{Y_2 - Y_1}$$

$$= k_I - \frac{m}{M} \cdot \frac{Y_f}{\Delta I}$$

which means that the employment multiplier will be less than the investment multiplier.

If labour-productivity falls, m will be negative and the employment-multiplier will be greater than the investment-multiplier.

If we want to study the ratio of the employment-multiplier to the investment-multiplier we may say that

$$\frac{K_e}{k_I} = 1 - \frac{1}{k_I} \cdot \frac{m}{M} \cdot \frac{Y_f}{\Delta I}$$

$$= 1 - \frac{\Delta I}{\Delta Y} \cdot \frac{m}{M} \cdot \frac{Y_f}{\Delta I}$$

$$= 1 - \frac{m}{M} \cdot \frac{Y_f}{\Delta Y}$$

$$= 1 - \frac{(m/M)}{(\Delta Y/Y_f)}$$

m/M is the rate of increase of labour-productivity and $\frac{\Delta Y}{Y_f}$ is roughly the rate of increase of income. Hence $\frac{m}{M} \div \frac{\Delta Y}{Y_f}$ may be called the *income-elasticity of labour-productivity*. Hence, if the income-elasticity of labour-productivity is less than zero, the employment-multiplier is greater than the investment-multiplier.

Another way of getting an idea of the relative value of the employment multiplier as against the investment multiplier is to assume that when income increases finally from Y_1 to Y_f, productivity increases from M_1 to M_f and employment increases from l_1 to l_f. Then

$$\frac{l_f}{l_1} = \frac{Y_f/M_f}{Y_1/M_1} = \frac{Y_f/Y_1}{M_f/M_1}$$

If $Y_f/Y_1 = 1{\cdot}2$ and $M_f/M_1 = 0{\cdot}8$, $l_f/l_1 = 1{\cdot}2/0{\cdot}8 = 1{\cdot}5$. This means that although income increases from 1 to 1·2 i.e., by 20%, employment increases from 1 to 1·5, i.e., by 50%. So, the employment-multiplier is likely to be 2·5 times the investment-multiplier.

As in the case of the investment-multiplier, time-lags can be introduced here also. If labour-productivity increases during the period in which the final employment level is reached, less labour will be required for producing the same real output and so the employment-multiplier will be reduced.

It may be mentioned again that though the investment-multiplier measures the multiplier-effect of investment on income, the *employment multiplier* does not indicate the multiplier-effect of employment on income. Instead, it is the ratio of the ultimate increase in employment as a result of autonomous initial increase in employment:

$$\frac{l_f - l_1}{l_2 - l_1}$$

In this sense, the use of the term 'employment-multiplier' is inconsistent in the face of the use of the term 'investment-multiplier', but the concept is useful.

One may add a final feeling that more attention needs to be given to (*i*) the employment-effect, and (*ii*) how it works out, because, as we develop, it is not merely the output but also employment—its volume, composition and nature—which is important: rather, at times, this is

more important to give the people a feeling of confidence which is born out of the state of 'I earn my own living'.

Determination of Multiplier

Since the multiplier concept is considered useful, a few words may be said about its evaluation (or determination) in practice, at least so far as the investment-multiplier is concerned.

In practice, the multiplier is taken to be the reciprocal of the marginal propensity to save, i.e., of $(1 - c)$ where c is the marginal propensity to consume.

Efforts have therefore been made to determine c by three methods:

(*i*) From a study of family budgets it is found out what amount of income is consumed by families with different incomes. The ratio of the increase in consumption to difference (increase) in income from family to family may be taken to be a measure of marginal propensity to consume.

(*ii*) From data for the nation as a whole, the consumption expenditure may be fitted to depend on (*a*) national income and (*b*) a trend factor:

$$C = a + cY + dt$$

where C = total national consumption; Y = national income; and t = time; and a, c and d are determined by the method of least squares. Then the value of c may be taken to be the marginal propensity to consume.

From the value of the marginal propensity to consume we may calculate the multiplier by the formula—

$$k = \frac{1}{1 - c}$$

A possible limitation of this method is that, in the relationship fitted, no account has been taken of the effect on consumption of income-distribution in society.

(*iii*) By fitting a relationship between national income (Y) and investment (I) plus exports (E)

$$Y = b(I + E) + e$$

We can find the value of b by the method of least squares.

Can the Multiplier be Zero or Negative

The multiplier gives the ratio of the increase in output (or income) in a new equilibrium position (as compared to the previous equilibrium

income) to the increase in demand (whether for consumption, investment, government expenditure or export).

In Diagram 20.1, starting from an equilibrium position P_1, if the demand rises, the BP_1C curve will shift up and take the position $B'P_2C_1$ or $B'P'_2C_2$. In either case, it cannot cut the OT line below P_1. Hence it may be argued that the multiplier cannot be negative though in a particular case it may be zero.

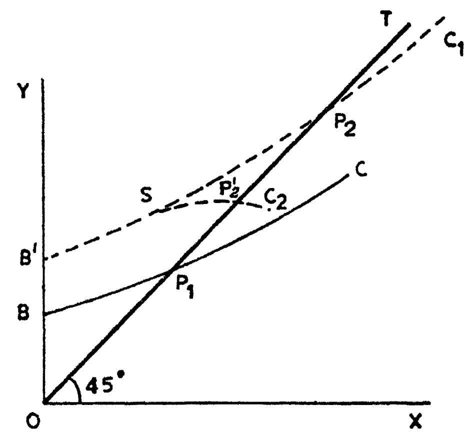


DIAGRAM 20.1

It is sometimes said that once an economy has reached a full employment stage—so that further expansion is not possible due to the shortage of natural resources or men—an increase in demand cannot lead to an increase in real income. Hence the multiplier (say, $\frac{\Delta Y}{\Delta I}$) will be zero due to no increase in income ($\Delta Y = 0$). In such a case the increase in demand, whether consumer's or producer's, is bound to be ignored. In actual life the full employment position is less likely to be reached as any persistent increase in demand is likely to result in researches and new techniques—to a reorientation of production, changes in forms and types of goods and even to a change in consumer's preferences to enable (more) production to satisfy more demand.

MULTIPLIER THEORY

We have so far seen how different multipliers are conceived or calculated and what useful role is attributable to a 'multiplier'. Let us now turn our attention to the 'why' and 'how' of the multiplier-effect.

A multiplier shows the effect (the multiplier-effect) of a change in a

causal factor on another factor (usually national income or employment). Human living is partly demonstrative and partly interdependent. An economic activity by *X*'s leads to some (or, similar) activity by *Y*'s. If one section (sector) of society begins taking ice-cream, others like t imitate it. If a certain proportion of us collects receipts through the process of illegal gratifications, i.e., through acts which are ethically or legally banned, some others also follow suit. If money-income becomes a status-symbol for some, others also accept it as their symbol and economic activities become increasingly money-oriented: work comes to be evaluated by and performed for money-returns. Hence a cause has a multiplier-effect if it increases in intensity through a demonstration-effect.

Again, the extent of interdependence of and the division of labour within economic activities lead to a multiplying influence of an initial cause. If *X* starts the production of a commodity, his act creates a demand for many goods and services: hence others start (or increase) economic activities to make these goods and services available. In doing so, they in turn increase their own demand for certain goods and services. The greater the interdependence and division of labour, the greater is likely to be the cumulative influence of a new economic activity.

Now, once a new phenomenon starts appearing in an increasing degree, it may be expected to go on and on, i.e., to go on expanding. But resources are limited and hence dampening influences set in. *X*'s cannot go on enjoying more and more goods and services daily: each enjoyment takes time and each *X* has only twenty-four hours to a day. Also, certain feelings (such as hunger) recur so that new time available for new enjoyments (and hence for new activities) gets limited. Similarly, *Y*'s cannot expand certain production activities for ever, due partly to limitation of resources. Raw materials may be in short supply: there may be a transport bottleneck: machinery may not be available limitlessly: and so on. Hence, either there may be a reversing of the phenomenon or an asympotic slowing down. In case there is a reversal, the multiplier effect has no (?) meaning. It is only when there is an asymptotic slowdown that one finds that a new level of *effect* is reached for all times—provided of course the cause continues to be present. (Diagram 20.2).

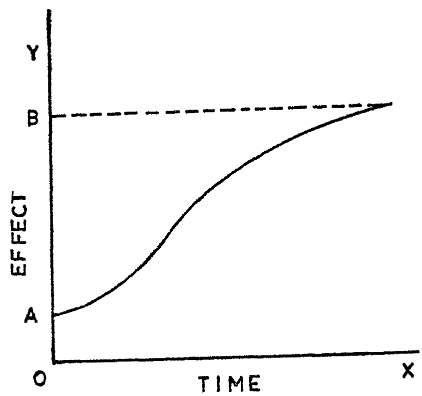


DIAGRAM 20.2

In the multiplier theory it is exposed how a new level of equilibrium of activities is reached: how an initial increase in investment leads to a permanent 'multiplied' increase in income. The table below reproduces an example of this effect, where we assume—

(1) that there is a recurring investment of Rs 100 per period,
(2) that half of an income is consumed constantly,
(3) that what is spent on consumption becomes the income of somebody for the next period,
(4) that what is saved accumulates (say) in the form of a bank balance, and
(5) that we ignore what the bank might create with that increasing balance in its coffers.

Period	*Investment*	*Income*	*Consumption*	*Savings*	*Accumulated saving*
1	100	100	50	50	50
2	100	100 + 50	75	75	50+75=125
3	100	100 + 75	87·5	87·5	212·5
4	100	100 + 87·5	93·75	93·75	306·25
5	100	100 + 93·75	96·88	96·88	403·13
			approx.		
6	100	100 + 96·88	98·44	98·44	501·57
7	100	100 + 98·44	99·22	99·22	600·79
8	100	199·22	99·61	99·61	700·40
9	100	199·61	99·80	99·80	800·20
Ultimately	100	200	100	100	

The table is self-explanatory or at least it can be understood with a little self-effort. Ultimately we find a stream of investment = 100, income = 200, i.e., the effect of investment of 1 leads to a stream of income of 2. Income increases rapidly during the first two periods: then the increase slows down. From increases like Rs 50 and Rs 25 we soon come in the seventh period and onwards to increases like 1·56 and 0·39.

If we change the assumptions, the picture given above would change. Thus if 80% of the income were consumed, the table would appear as on page 289.

It will be observed that the income crosses the 'double income' mark in the third year instead of practically in the ninth year previously. In the ninth year it crosses the 'quadrupled income' mark but it takes many more years—over fifteen years as against nine years in the first example—to reach its full effect.

However, to go back to the first example, the year to year investment in money must have been accompanied by (*i*) creation of additional capital, (*ii*) mobilisation of hitherto unutilised resources and manpower, and (*iii*) even a change in capital structure.

Period	*Investment*	*Income*	*Consumption*
1	100	100	80
2	100	180	144
3	100	244	195·2
4	100	293 2	236·2
5	100	336 2	269·0
6	100	369·0	295·2
7	100	395·2	316·2
8	100	416·2	333·0
9	100	433·0	346·4
10	100	446 4	357·1
11	100	457·1	365·7
12	100	465·7	372·6
13	100	472·6	378·1
14	100	478·1	382·5
15	100	482·5	386·0
Ultimately	100	500	400

Also, where does the money for continued yearly investment come from? Given a credit structure, there must be unutilised scope for creating credit or else the credit base would have been altered to enable the creation of credit for investment: and there must be willingness on the part of 'creators' to create credit. If the past 'created credits' are made good out of the accumulating savings then the balance of additional credit at the end of nine years will be 100. This is likely to be matched by an additional flow of real goods since made available. If savings are not thus made use of and if the amounts of saving continue to be in the income-earner's hand, it is likely to lead to a price-rise.

Again, why should the recurring additional annual investment of Rs 100 be made? It will be made if the entrepreneur, whether public or private, expects an increase in demand for the products and if the creators of credit have a faith in these expectations.

Consideration must also be given to the time-lags and periods of gestations in consumption and materialisation of production. The greater the time-lag, the slower the multiplier-effect is likely to be.

Finally, since we assume that investment leads to an increase in the creation of additional real goods—both producer-goods and consumer-goods—the consumption, and hence welfare, increases. If investment leads to an increase in capital goods without a simultaneous appropriate increase of consumer-goods, prices will increase and certain sections (i.e., those with relatively fixed income) will suffer in the transition period: this suffering may be acute and the transition period may be long depending on circumstances.

CHAPTER 21

The Accelerator

□□□□□□□□□□□□□□□□□□

Meaning

TO ACCELERATE means 'to make quicker' an action, e.g., the speed of a car, the velocity of a falling body or the chemical action in a solution. Acceleration is therefore the phenomenon of a thing becoming quicker. An Accelerator is the person or thing—a nerve in human body or an attachment in the car—that increases the action.

In mathematics, acceleration means 'rate of change of velocity'. If the distance covered by a moving particle is expressed as a function of time (say)

$$D = a + bt + ct^2$$

then acceleration is so defined as to be given by

$$\frac{d^2(D)}{dt^2} = 2c$$

If we wish we can give to $2c$ the name 'accelerator' and to the relationship the 'acceleration principle'.

If D is a more complex function of time (say)

$$D = a + bt + ct^2 + dt^3$$

then, $$\frac{d^2(D)}{dt^2} = 2c + 6dt$$

We may distinguish between $2c$ and $6dt$. The former may be said to be the fixed part of acceleration and the latter, the changing part. We

may, if we so like, call $2c$ an 'autonomous accelerator' and $6d$, a 'dynamic accelerator'. Other names can also be selected.

In economics, the word accelerator is used to indicate the increase in investment needed (expected to be made, or, planned to be made), for a unit increase in output (or, income or consumption). It is akin to—in essence the same as—the ratio of capital to output. Since capital (or capital goods) is valued at the capitalised value of output yielded by capital over its life-time, capital is greater in value than its output per unit of time. The accelerator is therefore greater than unity. It is sometimes also called the *acceleration coefficient.*

We can conceive of an average accelerator and a marginal accelerator. The ratio of total capital stock to total output may be called the average accelerator. The ratio of marginal increase in capital stock needed for marginal increase in output may be called the marginal accelerator.

Accelerator and Capital-output Ratio

In regard to the marginal accelerator concept, it is obvious that in the short period it will depend on the presence of unused capacity in the economy. Besides, the marginal investment will depend not only on technical considerations but also on the psychology of the investors or producers. Therefore it differs from what is called the 'capital-output ratio' in economics, which is purely a technical concept. Of course, on the whole, the accelerator must beat about the bush of the 'capital-output ratio': in the long run the former should tend to the latter.

For the student of statistics, we may try to approach the concept in another way. Assuming (*i*) a sustained 'autonomous change in investment', (*ii*) consumption to be function of current income, and (*iii*) a constant marginal propensity to consume, the multiplier shows the effect of investment on income. Similarly, we may conceive of the effect of an increase in income on investment, partly directly and partly *via* the consumption function and derived demand function for producer goods, including stocks and inventories. If

$$dY \div dI = k \text{ (multiplier)}$$

the student may write

$$dI \div dY = b \text{ (accelerator)}$$

He may look upon the multiplier as a measure of regression of income on investment; and upon the accelerator, as a measure of regression of investment on income. In one case investment is the cause: in the other, income is the cause. But the analogy would not be quite correct.

Investment I and income Y are not the same in both cases. In the case

of the multiplier there is an infinite time-difference between the initial investment (I) and the ultimate income increase. In the second the time difference is equal to the gestation period repuired for production-capacity to materialise.

Both the multiplier and the accelerator are greater than unity and so their rpoduct is not less than one as obtains in the case of linear regression coefficients.

In the first case, if one continues to invest '1' in each period, an increase of k in income will ultimately be enjoyed. In the case of the accelerator, if one wants an increase of '1' in income per period, an increase of b is needed in capital ultimately.

Even so, this statistical way of seeing things can help to grasp the concepts easily and understand the difference.

It may be added that, while the investment in the case of the multiplier is usually an 'autonomous' investment, the investment in the case of accelerator is termed the 'induced' investment.

Why an Accelerator in Economics

Why should b be called an *accelerator* or, as is also current, an *acceleration coefficient*? For it could easily be termed the marginal propensity to invest on the analogy of the marginal propensity to consume. In mathematics, acceleration relates to that by which velocity goes on increasing. There, an accelerating force, acting through velocity, increases the distance covered at an increasing rate. Similarly, if we come to realise that an increase in income causes an increase in capital stock which is many times the increase in income, we may, with *some* justification, use the term 'accelerator', borrowing it from mathematics.

Assuming full-capacity use of existing production facilities, an increase in income should mean an increase in the demand for consumption goods, and therefore an increase in the demand for capital in a certain multiple for 'production for sale' and another increase in demand for 'capital in order to maintain the conventional ratio of inventories to production'. Thus, if inventories are 10 per cent of annual sales (and hence of annual production by value, assuming prices to be constant), then if annual sales go up by 20%, the increase in the demand on producers will be (20 + 10% of 20%), i.e., 22%. If there is a need for three rupees worth of capital for a rupee worth of finished goods, then the demand for investment will go up by 3 times 22. Thus, investment is multiplied or (in some loose sense) accelerated. This is how one may justify the use of the word acceleration. This may not however appear justified if it is held that convention has led to the use of the words acceleration and accelerator in two different senses.

In practice, the accelerator is the ratio of investment to output, generally at the margin. Hence a controversy may arise to which we now turn.

Accelerator and Capital-output Ratio

If investment is held to be different from capital and if the former is measured in money, the investment-output ratio may be different from capital-output ratio. Now, the capital-output ratio is taken to be the ratio of capital stock in use to output: and in that sense it is said to be technical. In practice, when one says that the capital-output ratio is (say) 3, it is the ratio of money-values. So when it is said that, for purposes of measuring accelerator, investment is measured in money, while capital (of capital-output ratio) is not, one cannot attach much importance to it. Some maintain that investment has both an ex-ante and ex-post meaning while capital has only an ex-post meaning; this again is not of much value. Given an increase in income, and assuming full-capacity use of existing capital-stock, planned investment must be in the same ratio to income-increase which the existing capital-stock bears to the existing output. Investment may change if expectations about income-increase change, or if technology has changed since the last investment. Given an ex-ante investment, the ex-post investment should more or less be the same, unless the implementation part of investment is carried out with better or worse efficiency. The decision about investment is taken by the investor, and it involves

(*i*) his expectation of income or consumption increase,
(*ii*) his knowledge of the capital-output ratio, and
(*iii*) the extent of un-utilisation of existing capital capacity.

Hence, the accelerator is also subject to these forces. In addition, for expectations of decreases in income, the decrease in investment can at most be equal to the expense usually made on replacement. At the minimum it may be zero. It must be remembered that the capital cannot be contracted as rapidly as it is expanded. Its retirement depends on its durability, rate of obsolescence, etc. Hence, for decreases in income, the accelerator has an upper limit determined by the replacement expense, and a lower limit of zero.

Hence, it may be said that for positive increases in income the accelerator will have a value round about the capital-output ratio: conditions remaining the same, it cannot exceed it: it may be lower—even zero. In the reverse direction, for decreases in income the accelerator—or shall we call it the decelerator—may be zero at the minimum but cannot exceed a certain maximum set on the basis of replacement expense and decrease in output.

An interesting mathematical presentation in this connection is as follows:

Let gross investment (I_{gt}) at time t have 'replacement' (I_{rt}), 'autonomous' (I_{at}) and 'induced' (I_{it}) components:

$$I_{gt} = I_{rt} + I_{at} + I_{it}$$

and K_t be capital stock at time t so that

$$K_t = K_{t-1} + (I_{gt} - I_{rt})$$

$$\text{Also,}\quad K_t = K_{t-1} + \Delta K_{t-1}$$

$$\begin{aligned} \therefore\quad \Delta K_{t-1} &= I_{gt} - I_{rt} \\ &= I_{at} + I_{it} \\ &= I_{it} \end{aligned}$$

if we assume that autonomous investment is zero.

If CO_t stands for the capital-output ratio at time t, then, assuming full-capacity use, we can write

$$CO_t = K_t/Y_t \text{ and } CO_{t-1} = K_{t-1}/Y_{t-1}$$

and, if the two capital-output ratios be equal, the marginal capital-output ratio will be equal to them and to the ratio of increase in capital-stock to increase in income:

$$CO_t = CO_{t-1} = K_t/Y_t = K_{t-1}/Y_{t-1} = \Delta K_{t-1}/\Delta Y_{t-1}$$

Now, if induced investment is taken to be a linear proportion of the increase in income, then

$$I_{it} = g(Y_t - Y_{t-1}) = g.\Delta Y_{t-1} \quad (\textit{Hansen})$$

where g is the accelerator. Therefore,

$$\begin{aligned} g &= I_{it}/(Y_t - Y_{t-1}) \\ &= \Delta K_{t-1}/\Delta Y_{t-1}\,, \text{ if we assume that } I_{at} = 0 \\ &= CO_{t-1} \end{aligned}$$

But this result would not follow if we assume some other induced investment equation such as

(*i*) $\quad I_{it} = g\,(Y_{t-1} - Y_{t-2})$

(*ii*) $\quad I_{it} = g\,Y_{t-1} - K_{t-1} \quad$ (*Duesenberry*)

ACCELERATOR RELATIONSHIP

Let us derive an expression for the income(Y)-investment relationship, hereafter called the *acceleration relationship*. We know that the desired stock of capital depends on current income (*or* output) and rate of interest: expressed in a simple linear algebraic relationship it becomes

$$K = a + bY + cr,$$

where K stands for capital, Y for income, r for rate of interest, a, b, c

being constants. If K_0 be the existing capital-stock, then the desired increase in capital-stock is $K - K_0$. If it is to be reached in t years, then, $(K - K_0)/t$ will be needed this year. If we express K_0 in the same terms as K above, we can write that the investment will be given by

$$\frac{K - K_0}{t} = I = (1/t)\ b(Y - Y_0) + (1/t)\ c(r - r_0)$$

If rate of interest be assumed to be constant, we can, after suitable transformation of the origin and unit of measurement, rewrite it as

$$I_t = g\ (Y_t - Y_{t-1})$$

This hides within itself the facts that the accelerator effect, which is what g (called the *accelerator*) is said to measure, is subject to the adjustment factor — it may well be called the *reaction coefficient* — indicated by us by $1/t$, and to the changes in the rate of interest, r. Besides, we should not forget that in estimating 'expected sales' the businessman takes into account changes in income (or, consumption) in many past periods. So it is sometimes said that an investment decision involves *distributed lags*:

$$I = f\,[(Y_t - Y_{t-1}),\ (Y_{t-1} - Y_{t-2}), \ldots\ (Y_{t-\theta} - Y_{t-\theta-1})]$$

Really, there may be a number of factors that affect businessmen's expectations. Similarly, it may be argued that the effect of the rate of interest is not as simple as indicated by the relationship set out by us. Both the rate of interest (r) and the rate (or extent) of change (dr/dt) in it may affect investment: these may even affect $1/t$.

We may even look upon the accelerator b or g as determined or influenced by (*i*) partly technological requirements involving technique of production and technology, and (*ii*) partly by producers' psychology. There may be errors in calculating the technically necessary investment: there may even be factors influencing the minds of producers (producers who really constitute the means and media for investment decisions), making it pessimistic now and optimistic then: yet, in the long run, producers' decisions will be on the basis of technological requirements, i.e., according to what is called the capital-output ratio.

There is another way of looking at things. Let us start with the simple assumption that the value of capital must have a constant ratio to output so that we can write $K = gY$, which, on differentiation becomes $\frac{dK}{dt} = g.\frac{dY}{dt}$ or, in brief, $\dot{K} = g\dot{Y}$. But this does not take account of the lag in reaction of the interval between the change in demand and the new investment which it leads to. Further, it must have distributed lags to take account of changes in demand experienced in different periods in the past. Hence

$$I_t = f\,(Y_t,\ Y_{t-1},\ Y_{t-2}, \ldots\ Y_{t-\theta})$$

Or, starting from a normal capital-output relationship with a lag

$$K_t = gY_{t-\theta}$$

$$\text{or,} \quad K_{t+\theta} = gY_t$$

we may argue that the investment is proportional, not to the change in output, but to the deviation of actual output from normal output, taken to be (say) K_t/g. If b be the reaction coefficient (it may be taken to be the reciprocal of the number of years in which producers plan to make up the deficiency in investment), then the new investment shall be

$$I = \Delta K_{t+\theta} = b.g(Y_t - K_t/g) = b.\ (gY_t - K_t)$$

Here we take account not only of the time-lag but also over-capacity and under-capacity. It does not imply the assumption, which the simple accelerator concept does, that producers' investment is sufficient to keep up with the changes of demand period by period.

But if producers seek to minimise costs over time and increase (or build up) production capacity to meet the demand in the foreseeable future, they may equate the discounted marginal revenue and the discounted marginal cost over a 'planning period'. Assuming a secular rise in demand (at a steady rate) and taking price and production function as given, it has been shown that "when economies of scale exist, excess capacity will occur even with perfect forecasting"*; this is called *optimum over-capacity* and is a function of the discount rate, the planning period and the properties of the production function.

The type of investment function that has been indicated above takes account of the deviation from normal capacity: besides, it implies that entrepreneurs try "to balance capacity against output but do not invest (or disinvest) the whole amount necessary to do so in any one period"**: that producers aim at working at full capacity. But in order to take account of the over-capacity, we may use hK_t instead of K_t to indicate the optimum degree of utilisation of installed capacity:

$$I_{t+\theta} = b.\ (gY_t - hK_t)$$

Following this, Chenery formulated a *capacity-principle*

$$\frac{K_{t+\theta}}{K_{t+\theta-2}} = b.\left(\frac{Y_t}{K_t}\right) - bh$$

in addition to his *acceleration principle* in the form:

$$\frac{K_{t+\theta}}{K_{t+\theta-2}} = b.\left(\frac{\Delta Y_t}{Y_t}\right) + \text{constant}$$

* The inevitability of over-capacity in capital intensive industries was pointed out by Kuznets as long back as 1935 (vide *Economic Essays in Honour of W.C. Mitchell*, pp. 209-67).

** Not only Chenery (H.B.) but also Goodwin (R.M.) have in the past referred to such acceleration.

Chenery's study† once showed that for certain industries the capacity-principle is a better principle: for others, the acceleration principle is better.

So the simple acceleration principle, as the beginner knows it, is not as apt. Even in regard to what is said above, it should be added that the reaction coefficient, b, itself should be dependent on the nature of the deviation: for positive deviations, b should be low in value, partly because the need to expand is less urgent if large economies of scale make a low capacity factor (h) a normal feature. If the current output (Y_t) is less than normal output, then also it may be expected that b will have a lower value, partly because of technological limitations on a decrease in capital.

Theory of Investment

Let us now write out a number of mathematical relationships as representations of the theory of investment: in each the coefficient of income-increase will stand for the 'accelerator':

$$I = g(Y_t - Y_{t-1}) \text{ or } g(dY/dt)$$
$$I = g(Y_t - Y_{t-1}) - g_1 K_{t-1}$$
$$I = g(Y_t - Y_{t-1}) + I_0$$
$$I = g(Y_t - Y_{t-1}) + e$$
$$I = g(Y_t - Y_{t-1}) + d(r_t - r_{t-1})$$
$$I = g(Y_t - Y_{t-1}) + d(r_t - r_{t-1}) - W$$
$$I = g(Y_t - Y_{t-1}) + d(r_t - r_{t-1}) + e_t - W_t - K_{t-1} + I_0$$
$$I = g(Y_{t-1} - Y_{t-2})$$
$$I = g(Y_t - Y_{t-1}) + g'\,(Y_{t-1} - Y_{t-2})$$
$$I_t = b.g\,(Y_t - Y_{t-\theta})$$
$$I_{t+\theta} = b(K_{t+\theta} - K_t) = b(gY_{t+\theta} - K_t)$$
$$I_{t+\theta} = b(gY_{t+\theta} - hK_t)$$

Here K_t stands for capital stock; I_0, for initial investment; e, for the export of capital which the producers plan for; W, for the wage rate; b is the reaction coefficient, and h the capacity factor. The other symbols are obviously clear.

The justification for introducing W is that the higher the wage rate, other things remaining the same, the lower will be profit-opportunity and hence the lower the investment.

In the above formulae for investment we have not taken account of

† *Vide* Chenery: "Over-capacity and the Acceleration Principle", *Econometrica*, Jan. 1952.

price level. We can do so by dividing I, Y, r, and W by price P. This will give us another set of eight or nine equations.

By thinking out and quantifying many more factors that influence investment, one can write other (and more) relationships mathematically. In each there will be a common 'g times change in income'-factor and an accelerator shall loom up.

No matter how we conceive of the accelerator, it will beat about the capital-output ratio bush: it will be regressed on it, though there is not likely to be any regression-relationship between the two. The distribution of the accelerator about the capital-output ratio may well be stochastic: but it is difficult to assert this.

We may therefore think of the forces that affect the accelerator and give rise to it. Production necessitates the use of capital, for building which we have to consume less and work more; and it costs more to produce a capital-good than the value added by it to consumption goods produced with its help in a year (assuming that the life of the capital-good is more than a year). If we did not use capital-goods, there would be no accelerator effect. Since, for the accelerator-effect, there must be an initial income-change (or, expenditure-change), it can be brought about by drawing upon savings and hoardings and created money as well as cash-holdings. Cash and created money are pointers to a money-economy being favourable ground for accelerator effects. No doubt accelerator action will be influenced by demand for exports, invention of new techniques (meaning cheaper goods), discovery of new resources and even population growth.

Can the accelerator change with time, i.e., can it be dynamic? If the determinants of the accelerator change with time or with income or some other endogenous factor on the right hand side of the investment equation, then the accelerator will not be a constant but will change with time, i.e., it will be dynamic.

If the accelerator changes, one may study the effect of the change on production and consumption.

Multiplier and Accelerator

Between them, the consumption and the investment functions — or, shall we say, the multiplier and the accelerator — provide us with a basis for the study of the growth of income (*or* production). We can choose a consumption function, an investment function and together with the identity relation

$$Y = C + I, \textit{ or } = C + I + G, \textit{ or} = C + I + G + X - M$$

where C, I, G, X and M respectively stand for consumption, investment, government expenditure, exports and imports, work mathematically to

derive results for (*i*) income in equilibrium and (*ii*) income growth over time. We give below in equation form a number of such sets of relations and the results:

I. $$Y = C + I;\ C = cY + C_0;\ I = \text{constant } I_0$$
$$\therefore \quad Y = (I_0 + C_0)/(1 - c)$$

II. If the government levies a tax to invest the tax-revenue, t, then
$$Y = C + I$$
$$C = c(Y - t) + C_0$$
$$I = I_0 + t$$
$$\therefore \quad Y = \{(C_0 + I_0) + t\,(1 - c)\} \div (1 - c)$$
$$= t + \{(C_0 + I_0)/(1 - c)\}$$

III. If investment is taken to be a linear function of income so that
$$Y = C + I$$
$$C = cY + C_0$$
$$I = gY + I_0$$
$$\therefore \quad Y = (C_0 + I_0) \div (1 - c - g)$$

IV. If we introduce tax (t) as in II above, then
$$Y = (C_0 + t - ct) \div (1 - c - g)$$
$$= \frac{C_0}{1 - c - g} + \frac{t\,(1 - c)}{1 - c - g}$$

V. $C_t = (1 - s)\ Y_t;\ I_t = g\,(Y_t - Y_{t-1});\ Y_t = C_t + I_t$, which yields
$$Y^t = \frac{g}{g - s} \cdot Y_{t-1} = \left(\frac{g}{g - s}\right)^t \cdot Y_0$$
where Y_0 is the initial income and t indicates the time that has elapsed since then. This is called Harrod's growth model.

VI. $C_t = cY_{t-1};\ I_t = g\,(Y_{t-1} - Y_{t-2});\ Y_t = C_t + I_t$

This is known as Samuelson's Model and leads to the difference equation
$$Y_t = (c + g)\ Y_{t-1} - gY_{t-2}$$
which means that the rate of growth, G_t, is given by
$$G_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} = (c + g - 1) - g \cdot \frac{1}{1 + G_{t-1}} = (c - 1) + \frac{g}{\dfrac{1}{G_{t-1}} + 1}$$

where G stands for rate of growth of income. If G_{t-1} is positive, it has a depressing effect on the next-period growth-rate, G_t. But the greater the growth rate, the greater is the contribution of $[g/\{(1/G_{t-1}) + 1\}]$ to G_t, and hence the greater is G_t: this effect seems to be asymptotic, for greater is G_{t-1}, smaller is the

depressing effect $\frac{g}{1 + G_{t-1}}$. Ultimately G_t tends to $\frac{1-c}{g-1}$.

VII. $C_t = cY_{t-1}$; $I_t = g\,(C_t - C_{t-1}) = gc\,(Y_{t-1} - Y_{t-2})$; $Y_t = C_t + I_t + G$ which leads to

$$Y_t = c(1+g)\ Y_{t-1} - cgY_{t-2} + G$$

VIII. If in the Samuelson's model we use $Y_t = C_t + I_t + A$ instead of $Y_t = C_t + I_t$, and use the other two equations indicated there, we get what is called Hicks's Model. The resultant income equation is similar to that in the case of Hansen's Model:

$$Y_t = (c+g)\ Y_{t-1} - g\,(Y_{t-2}) + A$$

IX. $C_t = c_1 Y_{t-1} + c_2 Y_{t-2}$; $I_t = g\,(Y_{t-1} - Y_{t-2})$; $Y_t = C_t + I_t + A_t$ which yields the result

$$Y_t = (c_1 + g)\ Y_{t-1} - (c_2 - g)\ Y_{t-2} + A_t$$

X.
$$C_t = c_1 Y_{t-1} + c_2 Y_{t-2} + \dots + c_k Y_{t-k}$$
$$I_t = g_1 (Y_{t-1} - Y_{t-2}) + g_2 (Y_{t-2} - Y_{t-3}) + \dots + g_{k-1} (Y_{t-k+1} - Y_{t-k})$$
$$Y_t = C_t + I_t$$

This model takes account of investment and consumption with distributed lags.

XI. $C_t = cY_t$; $Y = C + I + A$; $\bar{K} = gY + at$, so that $\frac{d\bar{K}}{dt} = g\dot{Y} + a$; and $I = dK/dt = L$, or a or $-\alpha M$, according as actual capital stock at time t is less than, equal to or greater than $\bar{K}$, where

$\bar{K}$ = Desired stock of capital asset

M = Maintenance capital and αM is the portion of M which is not replenished

L = Net investment assumed constant from period to period

a = Trend value of capital

Hence, $$Y = cY + \frac{dK}{dt} + A$$

or, $$Y = (dK/dt + A) \div (1 - c)$$

If the capital is to increase at a steady rate of L units, then the gross investment must be $L + M$. At the minimum, even the maintenance effort may not be made, i.e., the gross investment may be zero and net investment may be $-M$; or the gross investment may partly cover M leaving a net investment of $-\alpha M$. Otherwise, the gross investment would be equal to $a + M$, where a is the trend value representing technological progress. So by putting L, a or $-\alpha M$ for dK/dt, we solve for Y.

This is called the Goodwin Model of trade cycles.

XII. $C = (1-s)Y;\ Y_d = C + I + A;\ \dfrac{dY}{dt} = -\lambda\,(Y - Y_d);$

$$\frac{dI}{dt} = -k\left[I - g\,\frac{dY}{dt}\right]$$

where Y_d is the total demand, which includes an autonomous expenditure A (on consumption and investment); where the response of output, Y, to desired demand Y_d is not instantaneous but lagged; and where the induced investment, I, is not only a constant proportion of the current rate of change of output (Y) but also has a continuously distributed lag of an exponential form and k is the rate of response. This is the Phillips Model.

It may be indicated, in this case, how the solution is attempted:

$$\begin{aligned}\frac{dY}{dt} &= -\lambda\,(Y - Y_d)\\ &= \lambda\,(Y_d - Y)\\ &= \lambda\,\{(1-s)\,Y + I + A - Y\}\\ &= \lambda\,(-sY + I + A)\end{aligned}$$

$\therefore\ I = \dfrac{1}{\lambda}\,\dfrac{dY}{dt} + sY - A,$ which on differentiation yields

$$\frac{dI}{dt} = \frac{1}{\lambda}\,\frac{d^2Y}{dt^2} + s\,\frac{dY}{dt}$$

Putting these values of I and dI/dt into the last equation of the model, we get a second degree differential equation in Y:

$$\frac{1}{\lambda}\cdot\frac{d^2Y}{dt^2} + \left(s + \frac{k}{\lambda} - kg\right).\ \frac{dY}{dt} + ksY = kA$$

We have given here a dozen examples of multiplier-accelerator models. Really, given a multiplier-relationship, an acceleration-relationship and an income identity, a model can be built up. If there be 10 multiplier-relationships, 12 acceleration-relations and 5 income-identity forms, one can build up as many as

$$10 \times 12 \times 5 = 600$$

multiplier-accelerator models.

Critique of Constants

Although it would be in some sense a repetition, it seems desirable to remind ourselves that ordinarily induced investment (I) is taken to be a constant proportion (or *relation*) of the change in demand (*or* consump-

tion *or* income *or* output). As once indicated by Giersch,* the accelerator abstracts from the following 'damping factors':

1. *Factor of expectations* — When the changes in the flow of demand (or sales) are unusually large and sudden, *g* will be smaller than when the changes are moderate and more likely to be long-term.

2. *Factor of finance* — When firms have to borrow, their investment activity may experience retardation due to

(*i*) the credit system not being completely elastic, through increases in the rate of interest

(*ii*) Kalecki's Principle of Increasing Risk on account of the deterioration in the ratio of owned capital to outside capital

(*iii*) imperfections of the capital market and the credit-control policy of banks

3. *Factor of Price* — When the price elasticity of supply is finite and the price elasticity of accelerated demand is greater than zero, the fluctuations in producers' purchase are damped.

These limitations appear to reduce the acceleration principle to a mere tendency, but it remains useful for business cycles theory to explain

(*i*) fluctuations in production and employment in industries far removed from consumption;

(*ii*) the upper turning point of the business cycle; and

(*iii*) the prior turning point at the maxima in the case of capital goods industries than in the consumers' goods industries. Besides the 'principle' is still used in connection with models for development and planning.

* Giersch: "The Acceleration Principle and the Propensity to Import", *International Economic Papers*, No. 4, pp. 201-02.

CHAPTER 22

Effects

Generalisations in Science

Generalisations or laws in a science consist of statements showing causal relationships between phenomena. Investigations in a science have the ultimate object of finding out the cause of a known effect or the effect of a known cause. Since cause precedes effect, logically if not temporally, all laws involve arrangement of events in a sequence. Knowledge of disconnected phenomena is no knowledge: if there has to be (scientific) knowledge the human mind should be able to conceive of the known events in a certain order. When such an order becomes available events appear to follow one another as cause and effect.

If a certain event is observed to follow another event on one occasion it is believed that it would follow that event on all other occasions. Such a belief is the extension of the general concept of the orderliness of Nature to cases of particular events. Nature is orderly because it does not change its designs; it exercises no free will. If there is a change it is due to the intervention of a (new) cause. Thus, the very notion of order in Nature is based on the principle of causation.

It is therefore sufficient to observe the sequence of two events on a single occasion to be able to connect them as cause and effect. Yet in practice it becomes necessary to make several observations and it is only when two given events are found to follow each other in a large number of cases that we conclude that they are causally related. This cleavage

between theory and practice is due to the fact of multiplicity of causes and intermixture of effects. Events are always found clustered together: groups of events (and not isolated events) follow one another, making it difficult for us to pair the right events in a sequential order. To be sure that there is correct sequence between any two events taken note of, we have to spread our observation over a wide area of occurrence. Let us repeat once more that all scientific generalisations consist in statements of causal relationship between phenomena.

Unnoticed Causal Relationships

The validity of our generalisations depends on our knowledge of cause and effect in their entirety. Logically it is sufficient to speak of *an effect* following *a cause* but, in reality, in this phenomenal world that we study, an effect is a collection, in a sense, of many events and so also is a cause. A collection of events called *A* might lead to events *B*, *C*, *D*, etc. Our mind, however, may fail to take notice of one or more of these and, consequently, regard only the remaining events as the full effect of *A*. All the events that together can be regarded as an effect might or might not belong to the same moment of time. In taking account of the complete effect of a cause it is possible to miss some of these events that coexist or follow one another. The difficulty of taking account of *all* the events that do in reality constitute the effect of a given phenomenon is great indeed. Where the events that escape our attention are not antagonistic to those that are known and taken account of, much damage is not done to the nature and character of our generalisations. But in certain cases the missed events are antagonistic to others and if and when they come to our notice our conclusions have to be suitably modified. Physics and economics are replete with examples of such unnoticed causal relationships.

Effects in Physics and Economics

There are many instances in physics where certain effects, unnoticed for a long time, were detected more or less accidentally by physicists during the course of their investigations. When a new, hitherto unknown, effect comes to light it is often associated with the name of the scientist who first detects it. To mention some of these we have in physics, Doppler effect, Cotton effect, Kerr effect, Purkinje effect, Joule-Thomson effect, Thomson effect, Piezo effect, Raman effect, Seebeck effect, Peltier effect, Becquerel effect, Barkausen effect, Stark effect, Zeeman effect, Villari effect, etc. Similarly, in economics we have Ricardo effect, Pigou

effect, Kalecki effect, Wicksell effect and Domar effect. The discovery of these effects have led to a fuller understanding of causal relationships between physical and social phenomena.

Our concern here is exclusively with effects in economics. The known effects are at times disputed on the ground that the newly detected effects, to a great extent, if not totally, neutralise the old and traditionally recognised effects. Let us take up for consideration the various effects one by one.

Ricardo Effect

Looked at from the physical side, production consists of acts of inputs leading to outputs. Thus, production is essentially a *process* occupying a period of time. Inputs relate to the present moment of time while the anticipated outputs relate to the future. As the main object of a producer is of maximise the excess (expected excess) of income over expenditure or, to put it in more rigorous language, as the main object is to maximise the discounted anticipated excess of income over cost, all the factors relevant to inputs and outputs have to be taken account of. Since the inputs relate to the present there is greater certainty about the calculated figure of cost than there is about the anticipated figure of income. As a matter of fact, all the inputs do not belong to the immediate present and yet, comparatively speaking, there is always greater certainty about the value of inputs than there is about the value of outputs.

For this reason there is an element of risk and uncertainty attached to an act of production. The memory of the past, the knowledge of the present and anticipation of the future, all play a part in production. Mainly due to the last factor on which production depends (though not entirely, as there can be error in recalling the past and evaluating its true import and in comprehending the total effects of the present), a producer is subject to disappointments and is led, therefore, to revise his plans frequently. Were such uncertainties not there economic activities would have been regular and producers' minds more at ease.

Production is, further, a lumpy and jerky act as it depends, and has of necessity to depend, on the use of capital. Because of this, there are time-lags in production activities — effects take time to follow the cause. Due to this interval of time between cause and effect, those forces get an opportunity to act which cause leakages in production activities. In short, we can say that due to the inability to foresee the future accurately, the operation of time-lags and the occurrence of leakages, production becomes a jerky process. These jerks are responsible for introducing those elements into the processes of production that cause trade cycles. Our object is not to explain the causes of the trade cycle or to suggest

remedies. It suffices for our purpose merely to note the broad fact that cycles in production activities are caused by the jerky nature of the course of production.

One broad fact about the trade cycle may be noted. There is an upswing followed by a downswing: that completes the cycle. During the upswing phase there is accelerated activity: production and consumption both go on increasing though not necessarily in the same manner. Because of increased and increasing production there is increasing investment, increasing employment of factors of production and, consequently, increasing payments to employed factors and increasing income of the hiring factors. Again, all these do not increase and cannot increase at the same rate. With the increase of consumption, savings of the people also increase but not in the same proportion. In short, during the upswing of the cycle all macro-entities register an absolute increase. The rates of increase do not however always register a rise. The reverse is the case during the downswing of the trade cycle. Economic activities slow down. Production and consumption begin to decline. Investment, employment and savings decrease in absolute terms. But here also these items do not decrease at the same rate. This dispartity between the absolute and relative amounts is due to the operation of time-lags and emergence of leakages. Trade cycles find in these their important, and in a way, their only cause.

The causes of the turning-points from the upswing to the downswing and then from the downswing to the upswing have attracted the attention of economists from time to time. Briefly, one could say that a turning-point is due to overshooting of the mark and the consequent necessity of retracing the steps. But this explanation can be spelt out in terms of economic concepts, more particularly in terms of savings and investment.

During the upswing of a cycle with increasing activity, wages, interest and profits tend to rise along with the rise of the prices of final products. Ricardo pointed out that, when prices rise, wages do not rise in the same proportion. This of course is due to the operation of time-lags in production activities. If and when wages lag behind prices and conceivably behind interest rates, labour as a factor of production becomes relatively cheaper, i.e., cheaper than capital. To the extent to which labour can be substituted for capital it would, therefore, tend to be employed in production in place of capital. In other words, the methods of production would become less roundabout. In the production of a given quantity of final goods there would be more labour and less capital employed. In the language of economics, there would be a change over to more direct or less roundabout methods of production. Now, if and when that change takes place, there is, to that extent, less demand for capital, less demand for borrowed money and hence less investment (or money). This is

called the Ricardo effect by Hayek. It is generally believed that the turning point comes when, due to overinvestment during the upper arc of the upswing phase of the cycle, the economic system is unable to sustain the rate of investment.

We have said that the trade cycle is caused by the operation of time-lags complicated by leakages. This is one way of looking at the phenomenon of cyclical fluctuations of economic activities. We can look at the same phenomenon from another angle. Due to time-lags in the adjustment of various economic forces, the balance between opposing forces is upset. One of these balances is that between savings and investment. During a boom (the upswing) investment increases. When it increases under the strain of time-lags the balance between savings and investment is upset. The situation thus resulting is described as overinvestment. The upper turning-point is then said to be caused by overinvestment which goes some way with declining activity, the downswing or depression.

The word overinvestment has, however, been used in two senses. Even when a small amount is invested it might become a case of overinvestment. Overinvestment means investment that is greater than what is consistent with some other entity in the economic system. Investment may be greater than what is justified by the demand for consumption goods; investment may be greater than savings. In both these cases one could say there is overinvestment. If people do not spend a sufficient amount on consumption goods and a part of the supply remains unsold, i.e., when the supply is in excess of demand, we can say that investment in the production of consumption goods is greater than what it should have been: it becomes a case of overinvestment. But when people do not spend a sufficient amount on consumption goods it means they save more. Hence, in this variety of overinvestment, investment is less than savings. In the other variety of overinvestment, investment would be in excess of savings. If and when such a situation is labelled overinvestment one could say that prices tend to rise and production thereby gets a fillip due to overinvestment. Such overinvestment becomes the cause of a boom. The first variety of overinvestment becomes the cause of the beginning and continuance of depression.

Hayek maintained that the upper turning-point is caused by under-investment rather than overinvestment. This he said on the basis of what he calls the Ricardo effect referred to above. During the upswing, prices of finished goods rise faster than the prices of services, particularly the price of labour. This is due to time-lags in the percolation of a shock to the bottom of the system. Due to this, real wages fall during the upswing and it becomes profitable for producers to substitute labour for capital wherever such a substitution is possible. The obstacles to such a substitution are both technical and non-technical in character. However, it can be admitted that some such substitution is always possible within limits.

To the extent to which production technique changes in favour of labour, the demand for capital decreases. Borrowing for purposes of financing roundabout methods of production slows down. Investment tends to decrease. If nothing else happens to neutralise this tendency there will be underinvestment.

But other things never remain the same. We, in economic discussions, often argue on the assumption of other things remaining unchanged. This assumption is made in order to enable us to concentrate on the effect of a given change undisturbed by other changes which follow or accompany it. Here too, when real wages fall (money wages lag behind prices of goods), investment decreases because labour is substituted for capital, i.e., early maturing projects or techniques are substituted for those that mature after a longer period. But one must remember that when real wages fall, profits (excess of income over cost) increase. And when profits increase production of goods always tends to expand. This leads to the expansion of the scale of production and, consequently, to increased investment. Hence, two opposing factors come into play: the increase of production leads to greater investment but the change over to less roundabout methods of production, to the extent to which that is possible, leads to decreased investment.

What the ultimate effect on investment would be depends on a variety of considerations. There is, for instance, the rate at which money can be borrowed; there is also the elasticity of credit and the elasticity of consumers' expenditure. While it is true that the Ricardo effect would operate to apply a brake to investment it is not the only effect of rising prices (falling real wages). It is most likely for the overall effect to be in the direction of increased investment.

It has, however, to be recalled that greater investment and overinvestment are not necessarily the same thing. When the producers are not able to sell all that they have produced for sale the unsold amount becomes an addition to investment or capital. The situation appears as one surcharged with investment—there is overinvestment. And when that happens investment made in the beginning of the production period is found to be less than the resulting saving. As compared to saving, investment is less and therefore it might also be called a case of underinvestment. However, this terminology is not applied to the discussion of the causes of the upper turning-point.

What then is the cause of the turning-point? Is it over- or under-investment? Since the word overinvestment is capable of dual interpretations it is best to avoid the use of this word and say simply that economic activity takes a turn downward when the actual sale of produce is found to be less than the anticipated sale. It is possible, however, for the down-swing to begin before this position is reached as it is also possible for it to come only when the situation deteriorates further. It is human psychology that determines the exact time of the turning point.

Pigou Effect

The Ricardo effect was discussed in relation to the trade cycle in general and the upper turning-point in particular. We shall discuss the Pigou effect in relation to the problem of employment in general and under-employment equilibrium in particular.

To produce a commodity we have to employ factors of production. Labour is not the only factor of production and yet when an economist talks of employment or unemployment he often has this one factor in his mind. Thus, when the Keynesians say that it is possible for an economic system to reach the position of equilibrium before the state of full employment is reached they have full employment of labour in mind. It is true that in the fuller discussion of the problem of employment all the productive resources are taken into consideration. But when a situation of underemployment is visualised attention is focussed on labour and other factors find their place only on the outer visual horizon.

With attention focussed on labour, what is the answer to the question: What is the effect of reduction of wages on employment? The classical economists believed that if there is unemployment wages would fall till all the labourers were employed. This of course took it for granted that wages were elastic, that is, there were no impediments to the reduction of wages. Two forces lend support to this view. First, if some labourers are unemployed they would express their willingness to work for a wage lower than the prevailing one. Second, if wages fall the cost of production would also fall and, therefore, production would increase on the assumption that the demand schedule for goods would not be adversely affected. Employment of labour would therefore increase. This analysis is correct as far as it goes and under the implied assumptions that other things remain unaffected.

This classical view about employment was thus based on the primary assumption that money wages were elastic and, second, that other things (say, the demand for goods) did not change. Keynes pointed out the fact that these assumptions were unrealistic. Money wages are sticky downwards and other things cannot remain the same. When wages in general fall, though the cost of production decreases, the demand for wage goods also falls. If it did not, the employment of labour would tend to go up. But with lower wages the total wage bill might be reduced and so also the demand for goods. The correct prescription that suggests itself, then, is that, to increase employment, wages should be raised.

What about under-employment equilibrium then? The analysis may now be directed into a slightly deviated channel. Suppose there is equilibrium with some labour force out of employment. If wages could be

lowered it would have the same effect on the production side as if the supply of money was increased. With this effect, the rate of interest would fall and, therefore, investment increase. Employment would thus go up. Will such a result inevitably follow? Can the money supply increase sufficiently to lead to a fall in the rate of interest? And with increased supply of money will the rate of interest necessarily fall to the needed extent? The first of these questions is difficult to answer. The second question permits a little discussion. The rate of interest might or might not fall with the increased supply of money depending upon the extent of preference of the people for liquid assets. If the preference for liquidity is quite high the rate of interest cannot fall much. The rate of interest and the rate of liquidity preference have, in equilibrium, to balance each other. Hence, if liquidity preference is high the rate of interest has also to be high. This is known in economic literature as the liquidity trap. The rate of interest is caught, as it were, in the liquidity trap. Again, it may be noted, even if the rate of interest fell the demand for funds for investment might not increase to the desired extent. When that happens we say that the demand for funds is not sufficiently interest-elastic. If these forces operate, employment cannot increase to bring about the state of full employment even when money wages are lowered. There would then be under-employment equilibrium of the economic system.

The stage is now set for a discussion of the Pigou effect. Pigou took account of the fact that people's consumption depends not only on their income but on their wealth also. If income increases, other things remaining the same, people would increase their consumption. If anything happens that has the same effect as increase of income, such as fall of prices, consumption must increase. In the reverse case consumption would decrease. But a change in price level does not merely alter the value of income; it also alters the value of wealth, i.e., of assets. If, as a result of fall of wages, prices fall to neutralise, at least partly, the good effect on production and employment, the value of assets held by consumers increases. The real value of cash balances rises when prices fall, notwithstanding the decrease of income. And when people realise that their wealth has increased in real terms they allow themselves to spend a part of it. Wealth is a right to future income. It is future income brought down to the present. Hence, when prices fall and the value of cash balances increases it is equivalent to an increase of future income. That is why people increase their consumption even when wages fall.

If, on the whole, consumption increases as a result of lowering of wages then producers would not suffer a loss consequent on increase of investment. Employment would in that case go up. The Pigou effect would conquer the normal effect of a cut in wages and Keynes's contention would be watered down. Yet, what the ultimate effect of wage reduction would be is hard to say. For, the end result is dependent on a variety of factors;

the initial change causes a series of further changes in the entire economic system. Reduction of wages will produce the immediate effect of increasing employment. The pay roll would change in one direction or the other according to elasticity of employment. With this change, or following this change, there would occur other changes in the rate of interest, the consumption and savings of the people and perhaps, to some extent, the technique of production. The Pigou effect would nevertheless operate and by its operation modify the rigour of other effects.

But it is believed that the Pigou effect itself needs a certain set of forces to enable it to operate in a macroeconomy. For instance, it has been pointed out that if the cash balances with some people are built up by borrowing money from private sources the increase of the value of wealth in real terms with some people is neutralised by the decrease of the value in the case of other people. Such a cancellation of effect does not, however, take place in the case of government debt as the government can be assumed to be indifferent to changes in the real value of its debt.

Effects Caused by Love of Inertia

An economic system loves to be inert: it dislikes change. That is why every change is opposed even though, at times, the opposition goes only to strengthen the initial change. Every action is followed by reaction. A mechanism and an organism react to every force that upsets their prevailing state. Where a static state does not obtain, an organism wants to maintain its prevailing dynamic state. In short, a change from the position that prevails at any moment of time is resisted. The force to counter the initial change is called the effect. The Ricardo effect and the Pigou effect both fall into this category.

Take the case of the demand for a commodity. When the price of the commodity rises the demand for it falls. This is how a consumer reacts to the change of price. The decrease of consumption is an attempt to preserve, as far as possible, the initial state. The cutting down of consumption is an effect. This effect is too obvious to escape the notice of consumers and producers. This change sparks off a series of changes. When the consumption of the commodity is reduced this change, in its turn, is disliked by the system and, consequently, the price tends to fall again. The initial rise of price ultimately produces the effect of a fall of price.

We know that when the price of a commodity is raised demand falls but this fall induces again a drop in the price. There is thus, first an autonomous rise in price and later an induced fall. As a net result, the final price is found to be located in between the initial price and the autonomously raised price. This happens in most cases, perhaps in all cases.

Invoking the Ricardo effect, we can expect investment in production to increase as a result of a fall of real wages. But this fall, in its turn, induces a decrease of investment. Eventually, then, investment must occupy a place between the original level and the higher level consequent on the fall of real wages. This argument abstracts from all other changes in the constituents of the complex economic system.

What is true of the Ricardo effect is true also of the Pigou effect. When money wages are lowered production immediately tends to increase and along with that unemployment decreases. But lower wages (unless elasticity of employment is quite high) result in decrease of consumption (at least of wage-goods). This is accompanied by a fall of prices which increases the real value of assets. This, as we observed earlier, is equivalent to higher future income. The consumers, therefore, become more liberal in spending money out of their current income. Consumption tends, thus, to increase. The final level of consumption must, therefore, be in between the original level and the level attained after the fall of wages.

Sigma Effect or Domar Effect

Sigma (σ) is a Greek letter that E. D. Domar uses in his growth model. R. F. Harrod's model and Domar's model are very similar. Both of them, in their growth model, start with a situation in which income is not constant, i.e., it is growing. And if income is growing it can increase at an increasing, constant or decreasing rate. And if the decreasing rate continues, eventually income would begin to decline. Growth and decline are thus two aspects of one and the same phenomenon. To come to the point again, Harrod and Domar are concerned with an economy in which income is growing and their object is the determination of the conditions in which income would grow without throwing the economic system out of equilibrium. For, if the system gets out of equilibrium, there is no predictable order in which it would fluctuate. Whether it is income or price or supply or demand, we are always concerned with equilibrium values. These are the only values that we can *determine.* They may not actually be found in an economy but they are in some sense the norm values which the erring economic order is striving to arrive at.

To start with, we must note the fact that a system is in equilibrium when the men behind the system are in equilibrium. And the men relevant for the study of a productive system are the producers, the investors. These people must be in equilibirum because they determine the magnitudes of economic variables. And when are they in equilibrium? To answer this question in the most general terms let us observe that they are in equilibrium when they are not disappointed, when they find that their anticipations have turned out to be correct.

A system is, then, in equilibrium when the producers are able to sell what they had planned to produce for sale. Output may be increasing or decreasing, but if it does so according to the plan of the producers the system is said to be in equilibrium. And when that happens, to use technical language, savings are equal to investment. Or, to express the same idea in other words, in equilibrium supply is equal to demand. That these two ways of saying the same thing are correct can be seen from the following identities. Savings are equal to total income minus expenditure, and investment is equal to total output (which equals income) minus what is sold off (which equals expenditure). Hence, when savings are equal to investment, what is bought (expenditure) is equal to what is sold. Hence, it matters little which expression we use to denote the fact of equilibrium of the system. Harrod, as we said above, equates savings to investment and Domar equates supply to demand.

The notations used by Domar are I for investment, Y for income, α for savings ratio (ratio of the amount saved to income) and σ for the ratio of output to capital.

The demand has to be equated to supply. So we take the demand to be equal to ΔY or $\Delta I/\alpha$, i.e., the demand generated by fresh investment. And supply is taken to be equal to σI, i.e., the supply generated by investment. One fact may be noted here. σ is the average ratio of output to investment or capital. In simple words, it is the average productivity of investment. We really need to determine marginal productivity so that we can equate it to the increment on the demand side. But in full employment assumed by Domar, and when small changes are considered, the marginal and the average productivities can be assumed to be equal.

Equating demand to supply we, therefore, get,

$$\frac{\Delta I}{\alpha} = \sigma I$$

What should then be the rate of growth of income to maintain the position of equilibrium and full employment? It is the value of $\Delta y/y$ under the condition that demand equals supply. Hence we get,

$$\frac{\Delta y}{y} = \frac{\dfrac{\Delta I}{\alpha}}{\dfrac{I}{\alpha}} = \frac{\Delta I}{I} = \alpha\sigma$$

This is the condition, then, of stable growth of income. Income must grow at the rate given by $\alpha\sigma$, i.e., the product of propensity to save and the productivity of investment. We know that Harrod's condition for equilibrium growth of income is that it should be equal to s/g. And Harrod's s is the same as Domar's α. And Harrod's g is the same as Domar's $1/\sigma$. Hence both Harrod and Domar arrive at the same result. All this, how-

ever, is conditional, i.e., it is true under the assumption made in regard to the behaviour of consumers and producers. And, further, this analysis abstracts from all other changes in the economic coefficients.

If the conditions of equilibrium growth are the same for both Harrod and Domar, why have they given rise to an *effect* in one case and not in the other? Why have we a Domar effect or sigma effect and not a Harrod effect or $1/g$ effect or simply a g effect?

Why Not a Harrod Effect?

There must be some significant difference between the g of Harrod and the α of Domar. If we do not speak of a Harrod effect, does it imply that what g stands for is not an indication of an effect produced by a change in some variable of the economic system? No, that is not the reason why we do not speak of a Harrod effect.

In the simple model of Harrod, g is the coefficient of the increment of national income, i.e., of the difference between the income of the current period and the income of the previous period. This increment multiplied by g is put equal to investment of the current period. It shows that producers invest an amount in production that is g times the increment of income. This is why we have called g the psychological coefficient in some of our writings. The value of g usually varies between 2 and 3. If, for example, income has increased by 100, producers invest g times 100. Investment is determined by two factors. First, there is the increment of income and, second, the coefficient g. This coefficient thus accelerates investment. It is not directly a technical coefficient. There are no doubt technical factors that influence the determination of this coefficient. But directly, it is a behaviouristic coefficient, if we could call it so. It shows the manner in which producers behave in regard to their investment. But Domar's sigma is a technical coefficient inasmuch as it purports to say that, when investment is increased, output is increased sigma times that amount. All measures are made in terms of constant money. This has come to be known as increased productive capacity. We know that the *multiplier* refers to the multiplied effect of investment on output or income. Here, in the case of the *accelerator*, we have the accelerated effect of income on investment as far as Harrod's model is concerned. In Domar's model, sigma does not show the accelerated effect of income-increase on investment. It shows rather the effect of investment on the capacity to produce goods. It is said, therefore, that investment has two types of influence on the economic system: it increases income (therefore demand) and it increases the productive capacity (therefore supply). This effect explicitly shown has led to the coining of the term, the sigma effect. Otherwise, this effect is also there in Harrod's model and in all models

that make use of acceleration. But whereas in Harrod's model the cause-effect relationship is from increased income to increased investment, in the case of Domar's model it is from increased investment to increased income or the objectively indicated increased capacity to produce.

Kalecki Effect

M. Kalecki said that investment is governed, among other things, by the principle of increasing risk. Whether we are concerned with the size of national income or with variations (cyclical or non-cyclical) in it we have to pay attention to the forces that determine the volume and rate of change of investment. Fundamentally, the same set of considerations go to determine the volume of investment of a macro-economy that determine the volume of investment by a micro unit. Yet, there is a difference in the result of the calculations based on those considerations.

As far as the supply side is concerned, the major factor on which income and employment depend is investment. The question that arises is on what does investment in normal circumstances depend? Investment is an input and the things produced are the outputs. A producer's aim is to maximise (the present worth of) the difference between income and expenditure. If production takes a certain period of time to complete its inputs and obtain its output, the producer behaves so as to maximise the present worth of the value of output, less the present worth of the value of inputs. This is a difficult calculation inasmuch as it involves foreseeing an uncertain future of varying length. It is comparatively easier to calculate the cost than it is to calculate the income. Anyway, a producer does make some kind of calculation about the possible net income and arrive at a decision about the investment to be made.

The cost of investment as such is the interest to be paid to the lender. The lender may be the producer himself, in which case the alternative use of money acts as interest. When the expected present worth of net income from production is greater than interest charges investment is considered profitable. From this it follows that investment would be carried to the point at which the rate of interest just equals the expected present worth of net income. Or, to use a somewhat simplified language that has come into use during the last few years, investment is pushed further and further till the marginal efficiency of investment becomes equal to the rate of interest.

This calculation becomes meaningful only when the rate of interest or the marginal efficiency of investment varies. If these magnitudes remain constant there is no limit to the extent to which investment can be carried. As far as the marginal efficiency is concerned, it depends on the cost of production and the value of goods turned out. As production

increases, the cost is likely to rise and the price of output to fall. But this result is to be expected only when production increases to a substantial extent. If the producer is one of many, an increase of his own output would make a very ineffective addition to the market supply of goods. The price of the product would, therefore, remain unchanged. Moreover, for the same reason, increase of his demand for the factors of production would make an insignificant addition to market demand. The price of factors of production would, therefore, remain more or less unchanged. If, then, the marginal efficiency of investment does not change when its volume is increased, the rate of interest must change if there has to be a finite limit to investment.

But the rate of interest depends on the supply and demand of loanable funds. This statement is made for the sake of simplicity, assuming that it does not cast a reflection on the other theories of interest. But, by his individual decision to borrow more or less, an investor hardly makes any change in the market demand for loanable funds. How then is the limit to be set to the magnitude of investment?

It is here that the Kalecki effect comes into the picture. Kalecki says that, due to uncertainties of various kinds, it is difficult to calculate exactly what the marginal efficiency of investment will be. There is, therefore, a significant element of risk in investing money in production. Since there is a possibility of loss, this loss will increase with investment. The investment of a producer is subject, therefore, to the principle of increasing risk. To keep the risk within limits, a producer does not go beyond a certain figure of investment. This figure depends on objective and subjective factors. On the objective side is the financial position of the producer and on the subjective his ability to make a forecast, his confidence in his own judgment and the strength or weakness of his spirit of adventure.

One thing needs to be stressed in this context. With increased investment there is not only greater risk of loss, there is also a better chance of gain. He who invests more does not therefore, merely take greater risk; he bears greater uncertainty. The investor is, thus, an entrepreneur who bears uncertainty. The limit to investment is, therefore, set by the capacity of the producer to bear uncertainty. Were this capacity unlimited there would have been no check on increasing investment by an individual producer.

But an individual is always a part of a bigger whole: he is a constituent element of a macroeconomy. Were all individuals to increase investment the total investment would be increased sufficiently to have its influence on the rate of interest. Moreover, with increased production financed by borrowing there would be an increase in the demand for productive factors sufficiently great to raise their prices. Thus, there is no escape from the limiting influence that economic variables exercise on investment.

CHAPTER 23

Ceilings and Floors

Meaning of Ceiling and Floor

A ROOM has three dimensions, length, breadth and height. Its height is bound on the upper and lower sides. The upper boundary is called the ceiling and the lower the floor. These are contrived to keep the contents of the room within limits. If the contents fluctuate these boundaries set upper and lower limits to fluctuations. Imagine an elastic ball (say, a golf ball) bouncing off from the floor of a room. It can go as high as the ceiling, after reaching which it must take a downward turn. On its downward journey it dashes against the floor and takes a turn again in the upward direction. Though not specifically meant for that purpose, ceilings and floors serve to reverse the direction of a moving object.

Function of Ceilings and Floors

The function of the ceiling and the floor of a room is to enclose it on the upper and lower sides. From another point of view, the purpose is to set limits to the height of the room. We, as economists, are not concerned with a brick and mortar structure. We study an economy which has, however, moving entities such as a room has. We took the case of a ball that bounces off from the floor and dashes itself against the ceiling of a room. An economy may be viewed as a room that has national income as its ball, travelling from its lower limit—the floor—to its

upper limit—the ceiling. National income is never steady nor does it travel in one and the same direction continuously. It goes up for some time and then begins to decline only to rise once again after reaching a certain limit at the lower end. While income thus fluctuates it might nevertheless tend gradually to rise. In other words, its up and down movements may be along a rising trend. Thus, there are two kinds of change; one is cyclical and the other a one-directional one.

Types of Cycle

When income goes up and down it is said to fluctuate in a cyclical order. The length of a cycle may vary from a few months to a long period of years. Those that are of a few months' duration are also known as seasonal cycles as they are caused by weather changes. Seasons are themselves cycles of climatic conditions and so income, to the extent to which it depends on season, must be cyclical in its variation.

Economists, following Schumpeter, talk of three types of cycle, namely, the Kitchin cycle, the Juglar cycle and the Kondratieff wave. These cycles have taken their names from those of economists who discovered them. The Kitchin cycle has a forty-month period, the Juglar cycle has a span of nine and a half years; while the Kondratieff wave extends over a period of fifty-six years. The present writer (J. K. Mehta) discovered a cycle of about three and a half years in certain grain prices in this country.

The difficulty of finding a cycle of any specific length is due to the fact that a number of cycles of different amplitudes occur concurrently. It is not impossible to separate one cycle from another to a certain extent; there are statistical methods of disentangling different cycles. But one cannot always be sure that the cycle one has discovered is *a* cycle or a mixture of two or more cycles. As a matter of fact, it may be that there is nothing like *a* cycle. In economic phenomena there is always intermixture of effects due to simultaneous operation of multiple causes. Due to such an intermixture one should not expect to find a cycle repeat itself in all its features. However, there is a general pattern that each cycle exhibits.

Endogenous and Exogenous Cycles

According to the laws of motion and the principle of uniformity observed in natural phenomena, if national income is increasing it should continue to increase. If it is decreasing it should continue to do so. Why are there, then, cycles of increasing and decreasing national income?

There can be two reasons for this. First, according to the above principles, if income can be conceived of as moving in a cyclical fashion to start with, it should continue to move that way. Second, the one-directional movement of income may be disturbed by external forces making it take turns. If and when cycles in income are accounted for by the first of these reasons they are called endogenous cycles. When they are due to the second they are called exogenous cycles.

In the case of endogenous cycles we can say that there is a built-in mechanism in the economic system due to which income generation is not steady. The economic system is unstable, like a radioactive substance; there is internal imbalance, disharmony of forces, due to which a continuous war is being waged with varying fortunes between opposing forces.

In the case of exogenous cycles we would say that an otherwise stable situation and a balanced economic order is disturbed by external forces. Economists have used the term erratic shocks for such external forces. These shocks are responsible for making income take turns. If we could widen the boundaries of our economic system so as to include the external universe, the whole system would become endogenous. The shocks would then be the manifestation of internal struggles between opposing forces. It may, however, be useful to keep the two systems separate in theory even if we realise that in practice it is perhaps not possible to do so.

Ceilings and Floors in Endogenous Cycles

Let us take a system that is endogenously unstable so that the income it generates goes up and down. What is the nature of the internal mechanism that causes the income to move in cycles? This question would need a long answer but we are not making here a study of the trade cycle so a brief statement should be sufficient for our purpose.

As we have observed elsewhere, the trade cycle is caused by lags and leakages in the working of an economic system. Adjustments to changes in some variables of a system are always delayed and, due to leakages, the force of some adjustments is wasted. We can also say that lags are implied in the nature of some coefficients of relevant variables. The multiplier-accelerator models are famous endogenous trade cycle models. However, the accelerator has to have a certain value if the income has to vary in a regular cyclical order. For other values of the accelerator the cyclical variations of income would be either damped or explosive (anti-damped).

When income increases either straight away or with ups and downs some explanation is needed to account for the turning point. The

explanation consists in the income encountering the *ceiling*. This idea of ceiling has become a commonplace of the economists' stock-in-trade since J.R. Hicks spoke of it. Others had, however, put forward the same views before Hicks. The ceiling is said to be reached with the full employment of resources. When resources in kind are fully employed further investment of money (induced by rising money incomes in general and rising real incomes for employers in particular) cannot lead to increased output. At this stage, therefore, economic activities must take a turn. Investment ultimately decreases and the income generation process is slowed down. The lower turning-point or the floor is reached when investment becomes zero and possibly negative in a sense, i.e., when capital goods are allowed to depreciate without repair or replacement.

One can say that the ceiling that accounts for the upper turning-point is a situation of scarcity of real resources while the floor that causes income to take an upward turn is a situation of abundance of real resources. One has to encounter two difficulties which may not be quite an obstacle to the acceptance of the above explanation of turning-points. The first difficulty relates to the concept of full employment while the second relates to the fact or the phenomenon of full employment of resources.

Full Employment Concept

We have just seen that, if and when all the real resources are fully employed, real income cannot increase any further. The validity of this statement depends on the meaning of the phrase full employment. All the resources cannot simultaneously be fully employed due to the fact that the technical coefficients of production are not perfectly elastic. When, therefore, one or more factors have been fully employed, the remaining factors are most likely to remain less than fully employed. The full employment ceiling is, therefore, encountered when all the technically employable units of all the factors are employed. It is in this sense that the term full employment must be understood.

A further concession has to be made to the concept of full employment. A situation is imaginable in which even when some units of all the factors are out of employment the full-employment ceiling is reached. That would happen when the unemployed units of at least one factor are such as cannot be profitably employed. To employ such units would be not only unprofitable for the employer but may at times lead to disguised unemployment. Conversely, it is possible that even when all the units of all the factors are engaged in work there may still be less than full employment. For, full employment does not necessarily mean the

employment of all the natural units (physical units) of factors of production; it means in reality the employment of all the efficiency units of factors of production. If a labourer is working for only one hour a day he is not fully employed. But this point need not be laboured further. Perhaps this difficulty does not block the way to the acceptance of the idea of full employment ceiling.

Full-Employment Ceiling in Growth Model?

The second difficulty referred to above relates to the phenomenon and not the concept of full employment. If resources are fully employed and cannot be over-fully employed so that income generation is retarded and growth is turned into decline can there be a rising trend of national income? If not, can the model be called a growth model? This question must be an embarrassing one to the builders of growth models. If the full employment ceiling has to live in a growth setting it must be treated as a temporary phenomenon. There must be a force in the system that pierces the ceiling, allowing it to rebuild itself at an extended height. What is that force? That force must be one that somehow leads to the increase of the available supply of factors of production or that which so changes the technique of production that, with the same supply of factors, a larger output can be obtained. This, however, involves a change in the quality or character of one particular factor, namely, organisation. It is really this factor, it is this force, that accounts for growth in spite of the system encountering a full-employment ceiling from time to time.

Here we might pose a pertinent question. If technique of production changes, if organisation changes (shall we say, its supply increases?) does it change due to endogenous or exogenous influences? If the former, we can say that the economic system is inherently unstable and that it has an explosive mechanism within itself. The system is then like an organism (as opposed to a mechanism) that possesses the needed growth potential. It is, then, an unstable system that has not attained the position of equilibrium, equilibrium in the sense of an overall stationary state—fluctuations round a horizontal trend.

Perhaps our economic systems are in reality of that nature. Or, who knows, there is an exogenous force acting on it from higher regions that leads to growth. In all models that find it necessary to commandeer the full-employment ceiling there must be a tacit assumption of the existence of some internal or external force that jerks the system, from time to time, and places it on a higher plane.

Bifurcation of the Ceiling

In the construction and examination of growth models two kinds of growth rate have been kept in mind. One is that which helps the system to keep labour supply fully employed; the other is that which is able to utilize fully the productive capacity of the system. In simple words, there is a growth rate which maintains full employment of labour and another which maintains full employment of capital. This is the bifurcation of full employment. Full employment should mean full employment of resources and not only full employment of one factor. But economists are not fussy about logical precision when it comes to using terms of everyday life. Thus, for them, full-employment rate of growth of income means the rate at which income must increase if the entire labour force has to remain employed. When it comes to the consideration of employment of capital they speak of full-capacity rate of growth of income.

Neither labour nor capital remains constant in an economy which is subject to forces of growth. The labour-force increases with the increase of population, sometimes due to economic prosperity and sometimes in spite of economic poverty. Capital-goods increase or the capacity of the productive system increases with the increase of investment. The two rates of growth are to a great extent independent of each other and, being so, it is not possible always to determine the rate of growth of income that would keep an increasing labour force fully employed and, at the same time, maintain full utilisation of productive capacity. Mention may here be made of the distinction between what has come to be called the natural rate, the warranted rate (full-capacity rate), and the actual rate of growth of income. Of these the last may be regarded as a compromise between the first two rates.

The Floor

When the ceiling becomes effective there is a check on the further growth of national income. When the floor becomes effective there is a check on the further decline of income. As we observed earlier, the floor is reached when investment ceases to decrease further. As shown by Harrod, when investment is increasing it appears to lag behind savings and when it is decreasing it appears to be in excess of savings. If the equilibrium rate of growth is defined by equality of investment and savings we should expect investment to stop decreasing if it has to be at par with savings. Net investment begins to decrease at the upper turning-point. Then it becomes zero: machinery and equipment are allowed to lie

unrepaired or unreplaced. This condition of production may be described as a case of disinvestment. And if this stage of negative investment is reached, or even before it, investment ceases to appear as being in excess of savings. The bottom is then reached and economic activity and all that goes with it begin to take a turn now in the upward direction. As we said earlier, while at the upper turning-point scarcity of factors of production is felt, at the lower turning-point abundance of factors of production is witnessed. Again, in other words, we can say that while at the upper turning-point there is abundance of production (consumption-goods), at the lower turning-point there is scarcity of goods. While during its journey to the upper point the economy registers rising prices, during its journey to the lower point the economy registers falling prices. Neither very high prices nor very low prices can be tolerated by an economy. And that is why it has to take a turn. But these are very vague statements inasmuch as the words very high prices and very low prices are imprecise. But as our object is not to explain the cause of the trade cycle it is not necessary to elaborate the above points.

However, one word might be said in this context. The turning-points are not determined merely by objective factors. Unless objective factors have their impact on the psychology of producers (and consumers) they cannot make either the ceiling or the floor effective. And that is why, at times when human psychology projects itself into the unborn future and brings expectations of coming events into the present picture, it expedites matters and consequently turning-points are reached earlier than otherwise. The upper turning-point is, therefore, not necessarily one at which the full-employment ceiling is reached nor the lower turning-point one at which the disinvestment floor is reached.

CHAPTER 24

Uncertainty

Uncertainty: Its Location in Economics

FROM THE point of view of economics a man is either a consumer or a producer. Every aspect of his life can be accommodated in consumption or production. When he is directly and immediately enjoying what he is doing he is said to be a consumer. When he is not enjoying what he is doing but expects to get some enjoyment ultimately or indirectly from it he is said to be a producer.

From one important point of view, therefore, the difference between consumption and production hinges on the element of time. Consumption has its location in the present while production, though itself located in the present, has its eyes, as it were, fixed on future events. That is why production is described as a time-consuming process. It is a *process*. Consumption by comparison is a *phenomenon*. Production necessitates looking forward. Uncertainty is, therefore, involved in acts of production. What has yet to come is not known with perfect certainty. Uncertainty attaches itself to acts of production: its location is in the production part of (economic) activities.

Uncertainty Caused by Spatial and Temporal Distances

There is no uncertainty about what one sees or senses unless of course

in cases where there is some defect in the sense organs or in the mind. Barring cases of illusion and delusion there is, therefore, no uncertainty about what is *here and now*. One can be uncertain only about what is at a distance, what is separated from one spatially or temporally. However, we shall concern ourselves only with cases where uncertainty is caused by temporal distances. What has yet to come and, therefore, yet to be experienced, has to be imagined, to be hunted out from the unborn future and dragged down into the living present. This involves a subtle mental process, the precise nature of which is not known and perhaps need not be known. Since production has its roots in the future and since uncertainty is attached to future events or experiences we shall concern ourselves with what we might call production-uncertainty, i.e., uncertainty arising from or associated with acts of production.

Nature of Uncertainty

As we have said above, uncertainty of the kind that we are concerned with is always about what does not exist, what is yet to be born. This, however, is a statement that can be challenged on philosophical grounds. The future is said to exist as much as the present. Future prices, supplies or incomes can, from the psychological point of view, be proved to *exist*. But we shall take the common, work-a-day point of view and say that the future has not yet come to exist. And since as producers we are interested in what is yet to happen we have to make a guess about its precise nature and magnitude. This we can do only if the future is (functionally!) related to the present and the past. If the future is born to the present we can guess the characteristics that it is likely to inherit from its parents. The events of this world are spread over time, each following from the preceding one. If we have to be able to guess the future from the present there must be a definite rule according to which the future follows from the present. This rule consists in the uniformity of Nature and the principle of causation. There is perhaps no very rigorous proof of the principle of causation (dependent as it is on the space-time continuum) but all our scientific knowledge rests on the assumption that such a principle of causation does govern all the phenomena of this phenomenal world.

Be that as it may, somehow our mind makes a guess about the future, about the outcome of decisions taken in the present. The future is not yet born but the human mind paints a picture of it. This picture differs from the pictures of events that have already taken place in an important respect. The former is a hazy picture as compared to the latter. The human mind is not like a machine; the vacillations to which it is subject blur the outlines of the pictures it paints for itself. In simple words, a

man is never quite certain about the outcome of his action. It is impossible to take account of all those variables of which the future outcome is a function. And this the human mind is conscious of. It does not know, however, how many and which variables have not been included in its calculations. Nor does it know whether the form of the function is a correct one. Sometimes the mind thinks one way and sometimes the other and this vacillation gives rise to a number of images of the future. These images are retained by the mind and it sees, as it were, several of them at the same time. If a producer has to take a decision in the present he must proceed on the basis of one particular image of the future. He has, therefore, somehow to select one image or construct a new one out of the many that are there before his mental eye.

Certainty out of Uncertainty

A producer can take no decision as long as there is uncertainty about the outcome of his venture. Assuming that all human actions are purposive and rational, there must be perfect certainty in the mind of the producer to enable him to take a definite course of action. It is, therefore, necessary to convert uncertainty into certainty for the purposes of production. If there are multiple images of the future, as there always are, they have to be converted somehow into one and only one image. For each possible course of action that a producer can adopt he has, thus, a definite, sharp image of its outcome. His final decision in regard to the course of action is determined by the profitability of the venture. That action which promises the maximum net return (in the ultimate analysis, psychic) is taken by the producer. We shall not dilate on this point further. Net return depends on cost and the present value of expected income. Discounting the future is a process that is partly objective and partly subjective. But this point is not germane to the understanding of the point under discussion.

Then Why Uncertainty Bearing and Why Profit?

If uncertainty is converted (because it is convertible) into certainty why should there be anybody to bear uncertainty? And if nobody has to bear uncertainty there should be no necessity for profit to exist as a reward for uncertainty bearing. But profit is there. It must, it would appear, be due to the fact that uncertainty does in some form and somewhere exist. If this is logically correct, two questions arise: In what form does uncertainty exist and who is the productive agent that bears (and therefore needs a reward for) uncertainty? To answer these questions

one must know the distinctive functions of the various factors of production.

Of the four factors (or agents) of production, the labourer supplies physical exertion, the organiser mental work, the capitalist waiting (he postpones consumption), and the entrepreneur supplies uncertainty bearing. Let us examine the contributions of the organiser and the entrepreneur to production. Production necessitates anticipating the future. It cannot be correctly foreseen for reasons that are well known. There is, therefore, uncertainty attached to every act of production. If uncertainty cannot be removed altogether (and it certainly cannot be) efforts must be made to minimise it. This effort consists in collecting all available information in regard to likely cost of production and likely income. The past and the present are treated as known factors and, in the light of the way in which one has behaved in the past and the other is behaving in the present, a forecast of the future is made. All these are complicated operations and require skilled and experienced men. These men are called organisers. Organisers are hired to do this work—those who do this work must be called organisers, no matter what they are known as in the business world. They reduce uncertainty to the minimum. They do not, however, bear this uncertainty: they are merely conscious of the fact that some uncertainty, from the objective point of view, remains. But, having done all that is possible, they place before the entrepreneur or, let us say, before the productive system, the final image of the future. This process of arriving at the final image may be taken to yield for the organiser or for the whole system a sure and certain picture of the future. It is in this sense that uncertainty is converted into certainty. Unless there is certainty, as we have emphasised earlier, no decision can ever be taken. The organiser has to feel or to persuade himself into feeling that, after doing all that was possible, uncertainty has been removed altogether. The work of production then begins, the labourer supplying physical work, the capitalist doing the necessary waiting (by supplying what he could otherwise have used as consumption goods) and the organiser now doing all the needed mental work. These three agents do all the active part of the work for the purpose of production. This is analysis. But a man may do two or three different types of work. He may supply both physical and mental exertion or either or both of these with the needed waiting.

While they thus begin to work, the ultimate uncertainty, which for the sake of production has been reduced to the minimum and then treated as certainty, remains for someone to bear. He who bears it is called the entrepreneur. He feels the pang of uncertainty but this is a passive phenomenon. The entrepreneur's function is a passive one. And if it were possible in actual practice to isolate this function from the rest, the entrepreneur would be a patient sufferer and a patient sufferer

only. But no man performs this function only, perhaps no man *can* perform this lone function. But ours is an analysis of the productive services.

Uncertainty and Profit

The word profit has been very loosely used in economics. In the business world it is used in the sense of excess of money income over money expenditure. But in economics it is the income of the entrepreneur who is an uncertainty bearer. Whoever bears uncertainty is an entrepreneur and whatever he earns as his remuneration for this sacrifice is called profit. The total earning of the man who goes by the name of enterpriser is not (pure) profit, it is a mixture of his other earnings for the work he might be doing as a labourer, an organiser, a capitalist and pure profit for uncertainty bearing.

Pure profit in this sense is an analytical concept and is, therefore, not of much use for the businessman. However, from what we have said it should be clear that in each and every case of production there is some uncertainty which somebody must come forward to bear and get rewarded for it. The entrepreneur can be considered as an agent that is hired by other productive agents to relieve them of the burden of bearing uncertainty. We are not explaining the theory of profit here, otherwise all the points involved in the study of uncertainty would be taken up.

One point may, however, be made clear here. The entrepreneur, like any other agent of production, always earns his full remuneration. Profit is, therefore, necessarily positive like wages, salaries and interest. Where the entrepreneur's income is less than the profit he had anticipated to get or where it is negative, it is because his total income is composed of profit (plus accidental gain or) less accidental loss. The position of the entrepreneur in the economic system is such as to make him appropriate all unexpected gains from productive activity and suffer all accidental losses. The possibility of such gains or losses are always there in all situations where uncertainty exists. And uncertainties do always exist, though every attempt is made to minimise them.

Alternative Projects and Uncertainty

The organiser's function (one of the legitimate functions) is to minimise uncertainty. He may not necessarily be the man who goes by the name of organiser in the business world. But there are various methods of producing a commodity and, more generally, there are various ways in which the investible money can be used for purposes of production. It is, therefore, the function of the organiser to ascertain the degree of

uncertainty attached to the outcome of each method of using money for the purpose of production. This possibility of alternative projects complicates the work of the organiser. With technical progress and scientific inventions the number of uses that can be made of capital increases and, along with this, the organiser's function becomes more difficult.

Suppose that, as far as one product (or one method of producing a commodity) is concerned, the income is Rs 1 lakh and its expectation is held with 95% certainty. Now, this is a mathematical way of indicating what the human mind thinks about the prospects of making a lakh of rupees. We have already said that some subtle processes go on within the mind to arrive at the figure of income. Objective and subjective factors play a part in the determination of this figure. But if uncertainty has been reduced to the minimum (in this case to 5%) the figure of one lakh of rupees is thereafter treated as a perfectly certain one. This follows from the fact that no decision can be taken as long as there is uncertainty. We repeat, the organiser (the man who bears uncertainty may also act as an organiser) has to make himself believe that the business is sure to earn a net income of one lakh of rupees.

When, therefore, the function of the organiser has been performed with care and efficiency no uncertainty remains for him. If there are many projects each has, then, its figure of income and all the figures are alike as far as the question of certainty goes. The problem thus becomes simplified, nay, it disappears. That project must be embarked upon which promises (in comparision to investment) the highest return.

Professor G.L.S. Shackle, whose work on uncertainty would be hard ever to excel, has after a detailed examination of the problems attached to the concept of uncertainty come more or less to the same conclusion. He rejects the mathematical computation of risks of uncertainty and says that some degree of surprise is attached to each outcome. After taking account of possible incomes and their appropriate degree of potential surprise the producer arrives at a conclusion about the project to be embarked upon.

The point that we find necessary to emphasise is that when uncertainty has been reduced to the minimum (which means in other words that when the mental picture of the future has been made as clear as it is possible to make it), the prospects of the enterprise must be treated as free of all uncertainties. The work can begin only then. And yet a cross-section of the picture would show that some ultimate uncertainty still remains. And this makes it necessary for someone to come forward to bear this risk. He, we repeat, does no active work: he is a passive bearer of uncertainty.

All Agents of Production Bear Uncertainty

Logically it is only the entrepreneur who bears uncertainty. But it is difficult to find any agent of production who does not perform the function of an entrepreneur as well. The most correct way of putting it is that all those who take part in production perform the functions together of the labourer, the organiser, the capitalist and the entrepreneur. The labourer who works for a promised wage cannot be quite sure that he will always get it as and when promised. His uncertainty is also due to the fact that the real value of money varies in an unpredictable manner. What is true of the labourer is true also of the organiser and the capitalist. Uncertainty is attached to every act of production due to the fact that production is a process extending itself into the future. Since no man can make a reliable forecast of future events some uncertainty always remains attached to production. This comes to be borne, in the system in which we live, by all the persons engaged in production. However, as we have explained earlier, except for him who undertakes to bear uncertainty, the element of uncertainty is reduced to zero by the effort of the organiser.

Is Uncertainty Preferred to Certainty?

Sometimes it is said that a consumer or a producer prefers uncertainty to certainty. Strictly speaking, this is a very incorrect statement. There is no man who would prefer a sum of Rs 1,000 with less than 100% probability of getting it than the same sum with 100% probability of getting it. It is maintained that there are gamblers who enjoy gambling. That is, however, quite a different matter. If there is such a gambler, who enjoys the very act of gambling, he will prefer what is cent per cent to what is less than cent per cent gambling. In other words, gambling too is desired more when it is certain than when it is uncertain. The activity of gambling, if it has to give maximum satisfaction to a gambler, must assure him that it is going to be a real gamble. The basic proposition, therefore, remains intact, namely, that certainty is always preferred to uncertainty.

It is said that, in certain cases, a man, when given a choice between (*i*) a sum of Rs 100 and (*ii*) a sum of Rs 200 with a probability, p, or a sum of only Rs 10 with a probability $(1-p)$, he would prefer the second choice to the first. The mathematical calculation that the man would carry on in his mind would be somewhat as follows: The prospect of getting Rs 100 is absolutely certain and so it is equal

to his actually getting Rs 100. If the probability attached to the sum of Rs 200 is 50% and that to Rs 10 is 50%, it is equivalent to his actually having got Rs 100 (50% of 200) plus Rs 5 (50% of 10). He would, therefore, prefer the second choice to the first. Such a case, however, does not prove that a man prefers uncertainty to certainty. The above example shows only one thing, namely, that, when it is possible to apply mathematical procedure to calculate utility, a man (a consumer or a producer) will prefer more utility to less utility. The Rs 200 which, for example, were expected with a probability of 50% were treated as Rs 100 expected with perfect certainty. The man converts, by the application of mathematical procedure, uncertainty into certainty. Whether a man who is confronted with a choice of the above kind does actually calculate in such a way is not the question with which we are concerned here. For our purpose, it is sufficient to note that uncertainty has to be converted into certainty before any decision or action can be taken.

Uncertainty and the Theory of Interest

We shall refer here to the Keynesian theory of interest. According to this theory, the rate of interest is determined by the demand and supply of money. Having decided what is to be included in money we would include in supply of money the total amount of it in the community or the economy. Demand for money consists of the amount that people want to keep with them uninvested (using that word in the ordinary sense). Demand for money thus means demand for liquidity. There are three purposes, as Keynes said, for which a man wants to hold money (demand money). First, he wants it for transaction purposes. Second, he wants it for precautionary purposes and, third, for speculation purposes. It is this third category of demand for money that varies even in the short-period. The speculation is about the rate of interest. The money that is held liquid and not invested is so held in the hope that the rate of interest might some day rise above the present level. There is thus a direct relation between interest rate and demand for money for speculation purposes.

The above is the skeleton of Keynes's demand and supply theory of money which has come to be known as the liquidity preference theory of interest. It will be clear that it is only because there is uncertainty about future events that the rate of interest that will rule in the market in the future is uncertain. And because it is uncertain no decision is immediately taken in regard to the investment of money. If it was known for certain that the rate of interest would be such and such a year hence, the money in the possession of a holder would be invested for a period of one year if the present rate of interest compared favourably with that a year

hence. What applies to the rate of interest a year hence applies also to more distant rates of interest.

The point we want to emphasise here is that the Keynesian theory of interest leans heavily on the phenomenon of uncertainty. If everything was certain the above explanation of the determination of the interest rate would not be available to us. In that case, it would become necessary to explore other regions for a satisfactory theory of interest. A more general theory that did not rely on uncertainty would perhaps conform very well with the supply and demand of savings theory.

CHAPTER 25

Disguised Unemployment

The Concept of Employment

THERE IS no question of employment or unemployment in the case of consumption. Only factors or agents can be said to be employed and there are no factors or agents of consumption. Production needs four factors that perform the functions of supplying physical exertion, mental exertion, waiting, and uncertainty bearing. Let us take the case of a labourer. A labourer is one who works with his body (to the extent to which he does mental work he is an organiser). When he enjoys this work, it does not constitute labour; he is not a labourer and what he does is not an act of production. It follows, therefore, that labour is not *in itself* enjoyable. Nevertheless, a labourer works because this disagreeable exertion promises a reward.

If production is, thus, an irksome activity its reward must be a desirable, enjoyable experience. Since there are only two opposites known to economists, namely, production and consumption, it naturally follows that the reward of production is consumption. In simple words, production is a sacrifice for the sake of some direct enjoyment.

We can, therefore, say that a man is said to be employed when he makes some sacrifice in the expectation of a return. This return may take any form: it may be money, anything possessing utility, a feeling of joy or a pleasurable experience. A labourer must be said to be employed when he expects a reward. And, to be logically most correct, the hope of a return or a reward is in itself a reward. This subtle point aside, one is

employed (and not unemployed) when one is making a sacrifice. Since nobody makes a sacrifice for nothing it is redundant to add that it should be in expectation of a reward.

Now, extending this analysis to factors of production in general, let us say that a person is employed when he works physically or mentally, waits (for the enjoyment of utility to be derived from consumption) or bears uncertainty. For, these are the only sacrifices that one can make. To the extent to which one enjoys work, waiting or bearing uncertainty one is not employed. Whether a factor is employed or not is, therefore, to be judged by the fact of sacrifice. It is not necessary to observe the final result of the sacrifices made. Whether a labourer is able to produce a commodity or whether a businessman is able to sell his produce matters little.

The Objective View of Employment

Taking a practical point of view some economists maintain that a labourer must be said to be employed only when he produces what he has been employed by the employer to produce. Going a step further, it is maintained that even if a labourer does his part in producing a commodity he cannot be considered to be employed if it does not add to the employer's income. According to this view it is not possible to say whether a labourer is employed or unemployed without waiting till the output is ready and sold. In other words, one can say whether a man *was* employed and not whether a man *is* employed. From what has been said in the previous section it can be easily seen that such a view of employment is logically not tenable. However, according to the present, practical point of view, a labourer may work sincerely and may even earn his wage and would yet be treated as unemployed if his contribution to the income of the employer turns out to be nil. Should it then be said that such a labourer is employed from his individual point of view but unemployed from the employer's point of view? And could we not carry this analysis a step further and say that a case is conceivable in which a labourer earns his full wage, makes an adequate contribution to the income of the employer, but, by so doing, decreases social welfare? From the social point of view is such a labourer unemployed? The moment one forgets the logically correct meaning of the word employment one is bound to be puzzled by a series of such questions.

Disguised Unemployment

Assuming that a labourer is unemployed when he makes no contribution

to the income of the employer, when is one to be said to be wearing the garb of employment? If a labourer is working as a labourer the *prima facie* assumption is that he is employed. But if he makes no contribution to the income of the employer he is, in effect, unemployed. His work, then, disguises his unemployment.

When a number of labourers are engaged in a productive enterprise, how are we to determine which labourers are unemployed? The answer is simple. Remove a labourer and see what difference it makes to output. If the output increases or remains the same as before we can say that the labourer was unemployed. That is, we have to judge by the marginal product of labour. But if this method is used to determine whether a labourer is employed or not every labourer would by turn appear to be unemployed. All that we can, then, say is that such and such a number of labourers are unemployed though they appear to be employed. There is disguised unemployment of so many labourers.

Now, if the marginal product of labour is zero, we can say that there is unemployment. But if the marginal product is negative what are we to say? As a matter of fact when the marginal product of labour is zero the workman cannot be called a labourer at all. For, a labourer is an agent of production and when there is no production he cannot be called a productive agent. Hence, when marginal productivity is zero the man engaged in work is not in the picture at all. However, in economic literature that does not care for logical niceties, such a man is treated as apparently employed but in effect unemployed. He provides a case of disguised unemployment.

When the marginal product of labour is negative the worker not only does not add anything to output, he draws something from it. He can, therefore, best be called a consumer. If we choose, we might say that he is a disguised consumer.

Disguised Employment

Just as there is disguised unemployment there can be disguised employment also. A workman, though appearing to be employed, is said to be unemployed if he adds nothing to output. Similarly, a man who is not actively engaged in producing anything may yet be treated as employed if, by being outside the enterprise, he helps to maintain the marginal output. The argument is simple enough. If a worker is to be called unemployed or, as we have suggested, if he is to be treated as a consumer, then, by removing him from active work, he becomes a producer. Such a disguised consumer reduces output and when he is thrown out of employment he increases output. He is, therefore, really employed though apparently unemployed. He provides the case of disguised

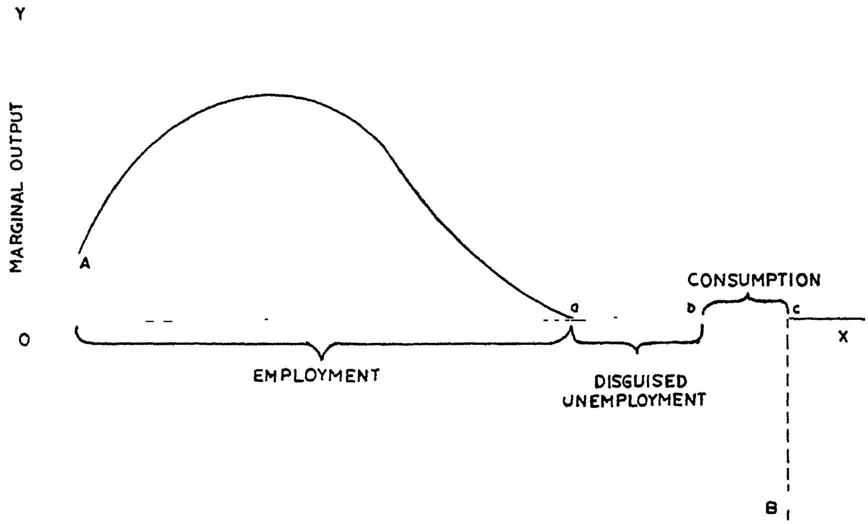


DIAGRAM 25.1

employment. Diagram 25.1 demonstrates these points.

On the vertical axis we measure marginal output and on the horizontal axis the number of workers engaged. Marginal output is positive till *Oa* workers are engaged in work. There is employment, therefore, of *Oa* labourers. Marginal output is zero over the range *ab*. There is, therefore, disguised unemployment of *ab* workers. Over the range *bc* marginal output is negative. There is, therefore, consumption instead of production. Workers over the range *bc* are not only unemployed, they are consumers. Were these workers to be dismissed and were they to remain out of work, they would help by so doing to raise the marginal output—it would become zero from a negative amount. They would, therefore, constitute a case of disguised employment.

INTER-INDUSTRIAL EMPLOYMENT AND UNEMPLOYMENT

Disguised unemployment in one industry or occupation may be disguised employment in another. A labourer, for example, may have zero marginal output in an industry where he is engaged in work but his marginal output in another industry, were he to be working there, might be negative. From the point of view of this industry, therefore, he can be considered as employed in disguise. If the quantitative relationship between labour and other factors is such as to result in the employment of labourers in every occupation to the point of zero marginal output, then every labourer individually is unemployed in disguise no matter

where he works and is employed in disguise in all the remaining occupations. But when the marginal output of labour in every occupation is positive there is no disguise either of employment or of unemployment.

CAUSE OF DISGUISED UNEMPLOYMENT

Disguised unemployment is due to marginal productivity being zero (where it is negative we have described it as a case consumption). For the cause of disguised unemployment we have, therefore, to search for the cause of marginal productivity being zero.

Marginal productivity can be zero only if it is a decreasing function of units of labour employed. And it is an ubiquitous technical phenomenon for the marginal product of a factor to decrease after it has increased over a certain range. There is always a technically ideal proportion in which factors of production combine for purposes of production. When the proportion of one factor to others exceeds that given by the ideal combination marginal productivity begins to decrease. But as far as it is possible to calculate and control the employment of a factor, its marginal productivity is not allowed to sink very low. Consequently, if it becomes zero the cause of this must be lack of knowledge or lack of control over employment.

In rural areas in countries where the industrial sector has not adequately developed and where, further, the people are illiterate, both the above causes of disguised unemployment are found to operate. As a result, a larger number of people continue to work on the fields than would be justified by considerations of net income. For one thing, it is difficult to know what precisely is the marginal productivity of labour in agriculture, especially when the members of large family are also working on their farms. For another, due to illiteracy and backward conditions prevailing in rural areas, coupled also with attachment to land, people do not readily move from their homes to industrial areas. This immobility, as we have said, is partly subjective and partly objective: it is due to ignorance and absence of a desire to move away from home and to lack of opportunities to get suitable employment, without much delay, in urban areas.

DISGUISED UNEMPLOYMENT OF CAPITAL

There are four agents of production, namely, labourers, organisers, capitalists and entrepreneurs, and these can be either employed or unemployed. Since employment is associated with production and since production is an activity involving sacrifice, there can be no

employment of non-human resources in any strictly logical sense. Yet, we do often speak of idle machines or idle material resources. In such cases, unemployed or idle resources would mean those material objects that are not actively used for production purposes. Machines that are not used in their normal way for the purpose for which they are meant or raw materials that are not used up at the rate at which they were intended to be used, constitute cases of unemployed capital goods.

Unemployment of capital goods in this sense occurs for various reasons. Once capital goods are bought they cannot be sold off the moment it is found that there is not sufficient work for them. Due to fluctuations in demand for final goods a producer has to use less or more capital goods than are in his possession. If he has to use less, some capital goods remain out of use. When he has to use more, the available capital goods are overused, i.e., used for longer hours every day.

But an employer has not much freedom in the matter of employment of capital goods. For, in most cases, when a machine is not used some labourers also have to remain idle. And an employer cannot with impunity play with the employment of labourers. It sometimes happens, therefore, that capital goods continue to be used when considerations of net income would not justify their use. In all such cases the phenomenon of disguised unemployment of capital goods is witnessed. As we have seen, much depends on the variability of the coefficients of production which is a function of objective (technical) and subjective factors.

Disguised Unemployment of Pure Capital

In simple, unsophisticated language pure capital means money or credit. Capital goods are bought with money (credit included). If all available money or pure capital were used up in buying capital goods there would be no possibility of unemployment of pure capital. But there is always in an economy a fund of investible money which is not actually invested but kept more or less liquid for reasons well known to economists. One could then ask whether the funds so kept can be (and if so when) unemployed. In the objective sense, all investible but uninvested funds are unemployed. Liquidity preference thus makes pure capital unemployed. But it is worth enquiring whether all money in liquid form is in every sense unemployed.

Money, as Keynes said, is kept liquid with three motives, viz., the transaction motive, the precautionary motive and the speculation motive. So far as that part of the liquid fund that is meant for transaction purposes is concerned, it should be easy to see that it is really employed in as much as it is used (during the transaction period) for making transactions. It is less easy to see that cash kept for precautionary

purposes is also employed. Cash reserved for speculation is employed in an important sense. It can, however, be said to be unemployed from a layman's point of view as it earns no interest in money terms.

The most important point to note in regard to liquid money is that it is objectively unemployed but not subjectively. If it is unemployed in the logical sense it should have no association with production. And if it has no association with production it must of necessity have association with consumption; it should become a consumption good. But money is a consumption good only if it gives direct enjoyment, if it serves as a miser's hoard. But liquid money meant for transaction purposes, precautionary purposes and for speculation is never a miser's hoard. It should, therefore, be regarded as employed.

It is, however, possible for a man to make mistakes and keep a larger sum of money liquid than is warranted by the consideration of net yield. The marginal yield of money in all uses should be equal, when the yield is calculated in terms of a common denominator. But due to immobility of money, accounted for by objective and subjective factors, marginal returns become unequal. It is, therefore, possible that some units of liquid money are yielding no returns. They can then be regarded as providing a case of disguised unemployment of pure capital.

CHAPTER 26

An Underdeveloped System

Introduction

AN UNDERDEVELOPED system may be conceived either with reference to potential capacity or with reference to expected target of consumption or even with reference to some target of labour employment. If we have in a country an unemployed ablebodied labour force with a standard of living which is not considered desirable, we may label its economic system 'underdeveloped' in so far as it desires to maintain its people at a reasonable standard of living. It may then be argued that other factors of production should flow to the economic system in order that the desire may be fulfilled. If these other factors refuse to flow in, the system may continue to be underdeveloped till it develops the other resources itself (which would imply that the other resources could be developed within the system even when they were sought from outside). One would feel like arguing in favour of a need for emigration of labour force in this situation and term the situation as one of surplus population.

The approach to the concept from the point of view of 'consumption' involves both the population situation and consumption aspirations. These aspirations are not determined merely by considerations of utility or necessities of life: they are also affected by what is called the demonstration effect. Under consumption aspirations, one may ask for a heaven on earth and call the situation one of permanent underdevelopment: yet the fact remains that consumption does have a hand in a judgment on underdevelopment.

Underdevelopment with reference to potential capacity implies a reference to both unused resources and inadequately (and improperly) used resources. It presumes that the system has the resources (including those which could be exchanged for needed resources) for more production and for more varieties of commodities. But more production of what? This is where 'consumption' comes in. If we have resources to produce what we want to consume and are not mobilising and utilising them, then a state of underdevelopment may be said to exist. It may be (as already indicated) that a part of resources have to be mobilised, processed and exchanged with other countries for other resources which we need. Given free trade, a society should be able to effect such an exchange subject to the theory of comparative costs. Given the spirit and the motive to help the relatively poor, developed economies can make a gift of certain required resources to an underdeveloped system: a developed economy may no doubt also render similar help out of political considerations as happened in some cases after World War II.

Certain Characteristics

There are certain other characteristics which tend to go with underdeveloped systems. Firstly, output whether per manhour, per worker or per inhabitant, is lower than that in developed systems. Secondly, there is a predominance of agriculture, the contribution of agriculture to national output is substantial and is generally more than fifty per cent. Thirdly, there is a high pressure of population on land: 'agriculture' and 'rural areas' have substantial unemployment and underemployment of human labour. Fourthly, there is more of illiteracy as also a lack of knowledge not only in regard to better use of resources but also in regard to more utilitarian consumption. Fifthly, there is a lack of capital. Sixthly, some existing legal and social institutions hinder development: the system of land tenure and inheritance laws as also cultural and religious patterns inhibit growth and lead to a lack of *motivation* to introduce better technique and improve productivity. Seventhly, there is a lack of urge to take risks and to expand: the people have a way of life which they *dislike* and yet do not have the urge to improve upon it through *initiative*, *enterprise* and *hard work*.

One might like to ask: Are these characteristics the results or the causes of an underdeveloped system? The first characteristic follows from our concept of underdevelopment. The second and the third features are more or less due to the fact that 'agricultural' and 'rural' states precede 'industrial' and 'urban' states in economic development and it is the latter which are aspired for. So are number four and six, but their persistence is a cause of lack of efforts to get out of the rut. Conse-

quently, there is a lack of supply of initiative, enterprise, organisation and capital, and this in turn becomes a cause of continued underdevelopment.

Desirability

Is an underdeveloped system desirable? Partly the answer depends on what the people living in the 'system' are consuming and what they want to consume. Mention has been made above of utilitarian consumption. In so far as consumption is not utilitarian and does not lead to a 'healthy mind within a healthy body', there will be ground for wanting a better state of consumption. But if this state has been reached and if more consumption is sought as a fashion—as a consequence of demonstration effects of the lives of other people—or under a belief that multiplicity of wants is conducive to greater welfare, then it is desirable that the state of underdevelopment continues. However, in general, an underdeveloped system has less than desirable consumption, though it partly has some undesirable consumption too. Thus, with undernutrition and malnutrition, the members of the system may be indulging in much drinking. Once the people have reached the state of desirable consumption, any element of underdevelopment (in the sense of availability of resources for use) is a source of strength and security against uncertainty and emergencies.

Besides, if the underdeveloped system is characterised by uneven distribution of income so that there is a gulf between the rich and the poor—technically, if the pyramid of income distribution is very broad-based and peaked—then it is easier to remedy the situation by producing more and redistributing the additional produce.

Of late there is another sense in which underdevelopment is said to arise and it is sought to remove it by seeking faster economic growth. Inadequacy of national defence or even international defence makes an economy appear underdeveloped. In the name of defending humans and humanity, there is an arms race, and 'more production', i.e., faster growth, is sought.

Remedy, at What Cost?

If a system is underdeveloped and if that state is considered undesirable, development, i.e., growth of output, should be sought in practice. But growth at what rate and at what cost? The sought-for growth will always necessitate a making of exertions and sacrifices to overcome various immobilities of materials and minds (ideas) as also to forego present (higher) consumption in order to build up capital (both material and

human). If the people are not prepared to pay this price or if (in a money economy) the growth efforts lead to more money chasing goods (like a rabbit chasing a tortoise), the costs of growth may prove to be unbearable to the people at large and particularly to people (of the middle class and the fixed-income group) who organise and manage the economy. It would then be desirable to seek a slower growth and a consumer goods oriented pattern of growth.

Let us use a little mathematics to throw light on the problem of development from a state of underdevelopment.

INCOME AND BUDGET DEFICIT

Assuming a multiplier, k, and a constant rate of tax on income, t, we know that

$$\Delta Y = k.\Delta G$$

where Y stands for national output consequent on recurring government expenditure, G.

If a change in the budget deficit (B_g) be defined as

$$\Delta B_g = \Delta G - \Delta T$$

where G stands for government expenditure and T for total tax, so that ΔG indicates an increase in government expenditure and ΔT an increase in tax revenue, then, since

$$T = t.Y = t(k.G) = tk.G$$
$$\Delta T = t.\Delta Y = t(k.\Delta G) = tk.\Delta G$$

we have

$$\Delta B_g = (1 - tk).\Delta G$$
$$\therefore \quad \frac{\Delta Y}{\Delta B_g} = k/(1 - tk)$$

But what is k? Assuming

$$Y = C + I + G$$

and $$C = c(Y - tY)$$

we get $$Y = c(Y - tY) + I + G$$

or $$(1 - c + ct)\, Y = I + G$$

$$\therefore \quad (1 - c + ct).\Delta Y = \Delta G$$

if I remains constant and government expenditure alone increases by ΔG. Hence,

$$k = 1/(1 - c + ct)$$

and $$tk = 1/(\frac{1-c}{t} + c)$$

Assuming that t is less than one (i.e., the whole of the income is not taken away by the government as tax), $1/t$ is greater than 1, so that,

$$(1-c)/t > (1-c)$$

and $$\{(1-c)/t\} + c > 1$$

Hence, $tk < 1$; $(1-tk) < 1$; $1/(1-tk) > 1$; and $k/(1-tk) > k$.

It can be concluded that the ratio of increase in income to increase in budget deficit is more than the multiplier (without deficit) assuming that the tax rate and the multiplier are constant and that additional expenditure is undertaken by the government alone. Also, and this is importtant, an increase in income is positively bound up with an increase in budget-deficit. With an increase in income, we may say, employment would increase: so, with a given population, a decrease in unemployment through government expenditure is bound to be accompanied by an increase in budget deficit as expressed by

$$\frac{\Delta Y}{\Delta B_g} = k/(1-tk), \text{ or, } \Delta B_g = \frac{1-tk}{k} \cdot \Delta Y$$

We would not be able to escape a certain rate of increase in budget deficit.

Income and Trade Deficit

Similarly, let us assume the presence of exports (X) and imports (M) in our income equation, exports being constant and imports being a certain ratio of income after tax:

$$M = m.Y(1-t)$$

Then, assuming other things as before, we have

$$Y = C + I + G + X - M$$
$$= cY(1-t) + I + G + X - mY(1-t)$$

$$\therefore \quad Y = \frac{1}{1-(c-m)(1-t)}(I + X + G)$$

If only government expenditure changes by ΔG, then

$$\Delta Y = \frac{1}{1-(c-m)(1-t)}\Delta G = k.\Delta G$$

so that k is the multiplier.

The balance of trade deficit. B_f, will be given by

$$B_f = M - X$$

$$\therefore \quad \Delta B_f = \Delta M - \Delta X = \Delta M \text{ (if } X \text{ does not change)}$$
$$= m(1-t).\Delta Y$$
$$= m(1-t).k.\Delta G$$

$$\therefore \quad \frac{\Delta Y}{\Delta B_f} = \frac{k.\Delta G}{m(1-t)\,k.\Delta G} = \frac{1}{m(1-t)}$$

which is positive since m is positive and $(1-t)$ is positive because t is less than one.

Hence, it may be concluded that with inelastic exports and imports proportional to disposable income, increased government expenditure for income-growth is bound up with increasing deficits in the balance of trade.

Incidentally, if $X = 0$, the above result can be deduced as below more conveniently:

$$\Delta B_f = \Delta M = m(1-t).\Delta Y$$

$$\therefore \quad \Delta Y/\Delta B_f = 1/m(1-t)$$

It can be concluded that an underdeveloped system which is managed for income growth only (or mainly) on the basis of increased public investment will find increasing budget deficits, as also increasing foreign trade deficits, unless the propensity to consume, the tax-rate or the import ratio (or, we must not forget, private investments and exports) changes. We have assumed here that technology does not change. We have also not taken account of monetary changes. We have done so to throw light on how mathematics can be used in regard to problems of growth of an underdeveloped system.

Budget Deficit and Tax Rate

If we think that a rise in deficit will increase prices and if an increase in price is not desired, we might like to decrease the budget deficit by increasing the tax rate, t. How will it all work out? We know that

$$\Delta B_g = (1 - tk).\Delta G = \text{(say) } b_g$$

and differentiating this with respect to t to study the rate of change of b_g with increase in taxation, we get

$$db_g/dt = -\left(k + t\frac{dk}{dt}\right).\Delta G$$

and since

$$k = 1/(1 - c + ct)$$

$$dk/dt = -ck^2$$

so that

$$db_g/dt = -k(1 - ctk).\Delta G$$

Now since tk has already been shown to be less than one and since c is less than one, ctk is also less than one (but greater than zero) and $1 - ctk$ is positive but less than one. Consequently, it can be concluded that the right-hand expression is negative: if taxation increases, then notwithstanding the effect on k, the budget deficit decreases. The multiplier-

effect will no doubt be smaller now: but the rate of change of income with budget deficit ($\Delta Y/\Delta B_g$) will rise because its differential coefficient with respect to tax-rate (t) will be given by

$$\begin{aligned}\frac{d}{dt}\left(\frac{\Delta Y}{\Delta B_g}\right) &= \frac{d}{dt}\left(\frac{k}{1-tk}\right)\\ &= \frac{d}{dt}\left(\frac{1}{1/k-t}\right)\\ &= \frac{-1}{(1/k-t)^2}\left(-\frac{1}{k^2}\frac{dk}{dt}-1\right)\\ &= k^2(1-c)/(1-tk)^2\end{aligned}$$

which is obviously positive as $1-c$ is greater than zero.

Hence, for a specific increase in income there will be less increase in the budget deficit.

Trade Deficit and Tax

A similar conclusion emerges if we investigate the effect of a higher tax on $\dfrac{\Delta Y}{\Delta B_f} = 1/m\,(1-t)$, for the differential coefficient with reference to t is $1/m(1-t)^2$ which is greater than zero.

Income Determinants

Or, again, if we simply express income (or, output) as made up of consumer goods (C) and producer goods (K) and then express C and K to be in turn dependent on a number of factors, we may write

$$\begin{aligned}Y &= C + K\\ C &= f(Y, a_1, a_2, a_3, \ldots)\\ K &= \phi(b_1, b_2, b_3, \ldots)\end{aligned}$$

where the a's and b's stand for different factors affecting C and K. Thus, a's may include tax, number of goods produced for consumption, income-distribution and credit facilities: and b's may include tax, rate of interest and rate of growth of income. Then, if we want to express a change in income in terms of these factors, we can write

$$\begin{aligned}dY = {} & (f_Y.dY + f_{a_1}da_1 + f_{a_2}.da_2 + \ldots)\\ & + (\phi_{b_1}.db_1 + \phi_{b_2}.db_2 + \phi_{b_3}.db_3 + \ldots)\end{aligned}$$

where (say) f_Y indicates the partial differential coefficient of C with respect to Y and (say) ϕ_{b_1}, that of K with respect to b_1.

Transferring the term containing dY to the left we can get the expres-

sion for a change in Y in terms of the effects of the different factors. For bringing about a rapid growth in Y we would be well advised to tackle those factors on the right side which have high coefficients associated with a change in them. Thus if f_{a_2} is greater than ϕ_{b_1}, then factor a_2 should be paid prior attention over factor b_1.

Income-growth and Savings

This is the distribution side. One could proceed from the production side as well. Thus, if g_l be the rate of growth of population as also labour force, and g_p be the rate of growth of labour productivity, then, assuming full employment of the labour force, no change in technology as also no shortage of complementary resources, the new income Y_1, as against an initial income Y_0, will be given by

$$\begin{aligned} Y_1 &= Y_0 (1 + g_l)(1 + g_p) \\ &= Y_0 (1 + g_l + g_p), \end{aligned}$$

ignoring the second order term $g_l g_p$. Hence, the rate of income-growth is given by

$$\frac{Y_1 - Y_0}{Y_0} = g_l + g_p$$

And, if the capital-output ratio be b:1, then, assuming full employment of capital, the required savings should be $b\,(g_l + g_p)$ per cent of income. This shows the importance of savings.

Again, given a target-rate of growth of income, g_y, and assuming no change in b, if the population-growth rate is high we shall be putting up with a lower increase—may be, even with a fall—in the productivity of labour: for, assuming that all increase is devoted to formation of capital, we have

$$g_y = b\,(g_l + g_p) \quad or \quad g_p = \frac{g_y}{b} - g_l$$

The experience of countries which have undergone development in the past supports the view that g_l depends on g_y: if g_y increases, g_l also increases. So, g_p is likely to be more affected by g_l. Hence in an employment-oriented plan, the target for income-growth must be carefully fixed so as not to promote complacency regarding productivity.

Employment and Agricultural Productivity

Yet another aspect that may be taken up with the help of mathematics relates to productivity in agriculture. An underdeveloped system is

generally a predominantly agricultural system. Not only does the agricultural sector account for about three-fourths of the labour-force but it also contributes more than half of the national income. The position of production in the agricultural and non-agricultural sectors may be illustrated by the following figures:

	Sector in which used		
	Agricultural	*Non-agricultural*	*Total*
Agricultural output	47	8	55
Non-agricultural output	6	39	45
			100

Let the agricultural sector have 70% of the labour force and 80% of population. If, now, one of the objectives is to decrease the dependence on agriculture, i.e., to reduce the labour force in agriculture, it would not imply a reduction in the agricultural output. If the labour force has to be reduced to 1/3 of what it is, then the productivity of labour in agriculture must show a trebling of what it is. If agricultural production is to be increased to a figure greater than 55, then the labour-productivity will have to increase further. This should involve an urge to work harder, an urge to expand business and an urge to experiment to overcome limitations in production. Besides, since one-third of the previous labour force must be put in possession of the necessary resources temporarily (in the short period) and made to pay for the same and own them in the long run, they must exercise thrift on the one hand and build-up the necessary institutions on the other.

General Considerations

An increase in agricultural production, particularly of food, need not necessarily be followed by an increased marketable surplus. Those who produce may consume. This will apply to the small cultivators and the labourers who prefer wages in kind. The surplus therefore has to come to a great extent out of the produce of the large cultivators. With a transfer of population to the non-agricultural sector consequent on the implementation of a policy of non-agricultural development, the demand for agricultural produce (particularly food) from that sector will increase. This is likely to put an upward pressure on prices: increasing prices may well give rise to speculation and hoarding for which bigger cultivators would have greater capacity. Hence, either a policy of compulsory procurement should be adopted or, particularly if it fails, those engaged in non-agricultural production must motivate and induce the agriculturists to possess, and hence to purchase, more non-agricultural goods.

Growth and Sociology

In a democratic set-up with adult franchise, it is likely that the government may not implement procurement policies strictly and may not impose an equitable burden of government expenditure on the agriculturists, particularly the bigger cultivators who may be more influential in mobilising votes. Hence, the non-agriculturists must—even in self-interest to escape greater burden and more privations—step out to motivate the agriculturist to break traditionalism, become more profit-minded and production-minded and consume more of what those in the non-agricultural sector enjoy.

Economists cannot be very helpful here with their suggestion concerning free market mechanisms. Sociologists must also lend a hand.

Leadership of the 'Wantless'

In an environment of large-scale production, induced by inventions and innovations, there is a danger of a growing gap between micro- and and macro-interest. Led by vested interests, and sometimes motivated by a sheer sense of frustration in reaction, non-agricultural and white-collar workers tend to work with lower efficiency on the one hand and demand (rather successfully) greater remuneration and security—in the face of inefficiency and indiscipline—on the other. The economist should not be asked to suggest remedies for this situation. hust as the maximisation method does not always help us to determine the equilibrium position, economics cannot be helpful here. A spirit of nationalism, sincerity, cooperation, sometimes even of efficiency and economy, cannot be bought for a price. Better education, better information, better communication of ideas, better social opinions—these would help but what is 'better'? The 'wantless' can intuitively interpret it or else History, that is yet to be written.

Underdevelopment and Developed Systems

Mention has been made of possible help by developed societies out of political or economic considerations. Is such help externally determined? One would suggest this is *built-in*. Even in Nature, plants growing adjacently have this social *give-and-take* of their resources. Human societies also, notwithstanding deliberate adverse endeavours, cannot long prevent the flow of facilities of their resources to the relatively underdeveloped systems under one plea or the other. Reason should guide developed systems to implement helpful policies *explicitly now* rather than wait till Nature forces them.

Name Index

Subject Index